A satellite-guided derivation of reservoir dam operation rules in ungauged locations.

A Case study in Vietnam

Conducted by:
Chiel Theodorus Eric Teunissen

Hanoi, Vietnam, 21/08/2024

UNIVERSITY OF TWENTE.

HUS VNU UNIVERSITY OF SCIENCE

Colophon

Title: Satellite-guided derivation of reservoir dam operation rules: A case study in Vietnam

Document: Bachelor Thesis CE.
Status: Concept version (no rights can be claimed from unverified, non-approved, documents)
Date: 21/08/2024

Author/Student: Chiel Teunissen
Student number: s2987260

Internal supervisor: Martijn Booij
External supervisor: Tien Giang Nguyen

Study: Bachelor of Civil Engineering & Management
Institute: University of Twente

Cover image: by Thương (2023)

Table of content

Preface

My name is Chiel Teunissen and I am a 4^{th} year Civil engineering bachelor student at the University of Twente. This thesis is performed at Vietnam National University of Science (VNU) in Hanoi.

I am very grateful for the opportunity I have been given to conduct my research at the VNU University of Science. My interest in modelling and hydrology and Civil engineering has driven me to search for a bachelor thesis assignment like this. The target audience of this thesis are fellow students and researchers in the hydrology field. Some background knowledge is required to fully understand the findings of this thesis. It is important to note that this report does not confer any rights or similar entitlements, and is intended solely for informational purposes.

For organizing this assignment, I would like to thank several people, starting with my supervisor from the University of Twente named Martijn Booij, who helped me find and carry out this assignment. Additionally, I would like to thank my external supervisor, Giang, for his excellent and pleasant support during the research, and I am grateful to him for the trust he has placed in me to contribute to his research. Finally, I would like to thank Mr. Hanh and Mr. Loi from HUS Office of Cooperation and Development for arranging practical matters regarding my time in Hanoi. And of course all other people who contributed to the process of conducting this bachelor thesis over sea.

Besides completing my thesis, this also signifies the end of my time as a bachelor's student. Achieving my bachelor's degree has not been without its challenges. Because I started my study in Delft and wanted to continue the same study in Twente. Many regulations, exemptions, and program adjustments were necessary to accomplish this with as less time loss as possible. Therefore, I would like to express my utmost gratitude to Judith Roos-Krabbenbos for enabling and facilitating a seamless transition from Delft University of Technology to the University of Twente. Of course, several other people have made this possible, such as the input from Peter, Gijs, and the examination committee, whom I also greatly thank for their trust in me to complete the program through a different path.

Summary

Lots of reservoir dams in the world are operated by operation rules. These operation rules are not always publicly available due to several reasons. When they are unknow, data based methods can be used to derive them although in-rsitu data is also not always available in regions lacking direct measurement capabilities. Deriving accurate operational rules for reservoirs remains a challenge for ungauged locations. Traditional methods relying on in-situ data are often inadequate and do have a low temporal resolution, prompting the need for alternative approaches. This study addresses this gap by proposing a methodology that combines remote sensing data, specifically from Sentinel-1 and Sentinel-2 satellites, with fuzzy logic techniques. The objective is to derive the operation rules of a reservoir dam based on readily available satellite data, thereby overcoming the limitations of data scarcity in ungauged basins.

The research method begins with the acquisition of satellite data, utilizing the high temporal resolution of Sentinel-1 and the high spatial resolution of Sentinel-2 to derive the Water Surface Area (WSA) of the reservoir behind the dam. This two satellite datasets are then processed by water volume curves to calculate the water volume. The integration of a water balance approach ensures accurate capturing of inflow and outflow dynamics, which forms the foundation for second model of this thesis, the fuzzy logic model. This model is used to obtain a certain set of operation rules of the dam and is calibrated using historical data obtained from satellites and validated against observed measurements, incorporating fuzzy rules to simulate reservoir outflows under various operational conditions. Additionally, the study examines how the model's accuracy can be determined in the absence of in-situ data.

The application of the proposed methodology to the Ban Chat reservoir shows an increase in the accuracy of reservoir modelling. The remote sensing-derived water volume time series achieves a Nash-Sutcliffe Efficiency (NSE) of 0.83, indicating a strong correlation with observed data and minor overestimation tendencies (1.93%). This underscores the robustness of using Sentinel satellites for generating reliable water volume estimates in data-scarce environments. Moreover, the fuzzy logic model shows small improvement over traditional demand curve methods, with an NSE of 0.59 and Mean Absolute Error (MAE) of 34.51 x 10^6 m^3, which are an increase of 7.27% an 1.48% respectively compared to the use of the long term average for modelling the outflow, suggesting its effectiveness in simulating reservoir outflows.

The findings highlight several strengths and limitations of the proposed methodology. While remote sensing proves effective in providing continuous data streams for reservoir management, challenges such as data accuracy and resolution persist. The study acknowledges the sensitivity of the fuzzy logic model to input variables and the need for further refinement in defining membership functions to enhance model accuracy. Moreover, the reliance on derived measurements for validation underscores the importance of improving validation methods through complementary approaches like altimetry satellites.

In conclusion, this research demonstrates the potential of integrating remote sensing data with fuzzy logic modelling to derive operational rules for reservoirs in ungauged locations. By leveraging the strengths of Sentinel-1 and Sentinel-2 satellites, the study establishes a framework for managing reservoirs where traditional data sources are limited. The calibrated fuzzy logic model shows promising results in simulating reservoir outflows, providing insights into effective water management strategies. However, the study also identifies several limitations, including data availability and model sensitivity, which require attention in future research efforts.

To enhance the applicability of the methodology, future research should focus on refining input variables and membership functions in the fuzzy logic model. Additionally, exploring alternative validation methods, such as integrating altimetry data with remote sensing outputs, could improve the

accuracy of reservoir management models. Moreover, expanding the study to different geographical regions and reservoir types would validate the generalizability of the proposed approach. Finally, incorporating hydrological models to predict inflows more accurately could further refine operational rules for reservoirs.

List of figures and tables

List of Figures:

List of Tables:

List of abbreviations

ABBREVIATION	FULL EXPANDED MEANING
BC-DAM	Ban Chat reservoir dam
DB	Decibels
DEM	Digital Elevation Model
GEE	Google Earth Engine
HUS	Hanoi University of Science
IN-SITU DATA	In-sitution data. E.g. data which is collected on site
KPI	Key Performance Indicator
MAE	Mean average error
MF	Membership Function
NDWI	Normalized Difference Water Index
NIR	Near Infra-Red
NSE	Nash sufficient efficiency
OBS	Observed
PBIAS	Percentile bias
RMSE	Root mean squared error
ROI	Region Of Interest
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SIM	Simulation
VH	Vertical-Horizontal
VIC	Variable Infiltration Capacity
VV	Vertical-Vertical
WL	Water Level
WSA	Water Surface Area
WSM	Water Surface Mapping
WV	Water Volume

1. Introduction

This chapter gives information on why this research is performed. Done by giving information on the background of Vietnam with its water security issues. Followed up by the current available knowledge on the solutions for these water security issues resulting in the research gap which is encountered in this research.

1.1 Background

Water as resource is one of the most important resources of everyday life. It´s used for different kind of purposes such as industrial and agricultural use, recreation and navigation. Although water could also be a threat to humanity explained as water security. The United Nations Water Commission defines water security as “the ability of a community to maintain sustainable access to sufficient quantities of water of acceptable quality to sustain livelihoods, human health, and social-economic development, ensuring against waterborne pollution and water-related disasters, conserving ecosystems in an environment of peace and political stability.” The report “Vietnam’s Water Resources: Current Status, Challenges, and Security Perspective” states that the score of Vietnams water security needs to improved much. To do so, “*The management of water resources in Vietnam needs to focus on addressing issues such as controlling the exploitation and use of water, protecting water quantity and quality, and ensuring national water security*.”.

Vietnam is one of the countries experiencing heightened impacts of climate change due to its location in multiple river deltas combined with a tropical climate. According to Pham et al. (2023), the average annual discharge of most river basins in Vietnam tends to increase by 2-18% per year. However, this increase in discharge is primarily concentrated in the wet season, while during the dry season, the discharge in the driest month tends to decrease, exacerbating water shortages and scarcity even further. Additionally, it is important to note that 60% of all water in the rivers of Vietnam originates from upstream countries (Lai et al. (2009)). For the *Red River*, located in the north of Vietnam, this percentage is 40% (Vu Hong Chau, n.d.) making Vietnam very dependent on China for their water supply.

Improving the water security can be done in several ways and so, several study topics are proposed by the UN. One of them, “improving monitoring data from outside the country using advanced technology” is used as motivation for this thesis.

Modelling the streamflow in the upper part of the *Red River* can contribute to improving the water security. Although a large amount of data is needed to build those models. In the past, a lot of data was collected and shared between China and Vietnam. However, this is no longer the case since the 80s. In addition, collecting in-situ data is becoming more difficult and expensive, making the data scarce according to Kebede et al (2020).

Nevertheless, data collection is essential to develop a streamflow model of the river. The Red River is influenced by many reservoir dams, build in the past decades. These dams changes the natural flow patterns of the river based on their operation rules. The operation rules of these dams are often unknown, making it hard to model the streamflow in the river. Obtaining them can be done in various ways. The easiest way is improving agreements between countries, which is still in its infancy (Tone, 2023). However, this approach is not desirable since there are still implications on costs and difficulties. An independent form of data collection, independent of location, is necessary to build a model capable of predicting the behaviour of a reservoir dam in the *Red River* autonomously. Besides this, China could also benefit from defining the discharge of their rives in other methods. A potential solution is the utilization of remote sensing data openly available. Spatial data can be used to obtain information needed to build a model of the *Red River*.

1.2 State of the art

Improvements can be done by modelling the catchment area together with the upstream part of the rivers. Modelling this is already widely done over the world and is currently under progress at the *Hanoi University of Science (HUS)*. Modelling the overall river gives a need in the modelling behaviour of reservoir dams located in the river. Modelling dams is also widely done already although, there are lots of different possibilities in doing so based on the exact location and available data(Wang et al., 2010) (Nikoo et al., 2013) (Mateo et al., n.d.). Mapping nearly all water systems is needed although this thesis only focuses on the *Red River* delta. So when river or basin is said, the *Red River basin/delta* is meant. This chapter gives more insight in the current state of the art and the goal of this thesis.

1.2.1 Flow forecasting model

Increasing the knowledge of the water flow requires mapping of whole flow regimes in Vietnam and upstream countries. Modelling the flow in the river catchments can be done by several methods, such as Soil and Water Assessment Tool, Hydrologic Engineering Centre's Hydrologic Modelling System as well as a Variable Infiltration Capacity Model (VIC) and much more. Barely all model requires lot of data as input. Currently the HUS is working on a VIC-model, simulating the flow of the *Red River*. This model can be split up in three separate sub models defined as the VIC-Runoff, VIC-Reservoir and VIC-routing model. The model requires data sets about precipitation, evaporation, land-use, Temperature, soil types etc.

Moreover, in the last few decades, several changes have been observed in the flow regime of the *Red River.* The construction of dams in the upper part of the *Red River* basin have a major impact on the flow regime of the *Red River* in Vietnam (Winkel, 2022). *Red River* dams are causing repeated floods as well as droughts in Vietnam, according to L. Pham(2021). The Hanoi University of Science (HUS) is already in possession of a working VIC-model for different sections of the *Red River* basin. Although running the VIC-model over the whole *Red River* gives poor results which are caused by the lack of information on the behaviour of the reservoir dams, on which remote sensing can play a key role in defining the behaviour of the dams.

1.2.2 Remote sensing data for hydrological purposes

Remote sensing, as defined by NASA (2022), involves obtaining information about an object from a distance. In March 2024, there are around 9494 active satellites around the world of which 1052 are used for Earth Observations. This number is still even increasing which provides more and more information and possibilities in earth observations (Ieva, 2023).

According to Victor (2023), remote sensing enables the monitoring of water bodies, precipitation, and snow cover, providing valuable insights into hydrological processes. Its capability to detect changes in water levels and identify potential flood risks enhances water resource management and flood forecasting. NASA (2022) highlights some advantages of remote sensing relevant to this thesis, including providing information where ground-based measurements are unavailable and enabling continuous monitoring of our planet.

Lots of studies have been done before on this topic but they directly show the relevance of location specific properties. Remote sensing, particularly satellite observations, can be a valuable tool in defining the operation rules of a reservoir dam when in-situ data is unavailable. Eldardiry and Hossain (2019) and Ali and Sridhar (2019) both demonstrate the use of satellite data to estimate reservoir storage change and outflow discharge, with Eldardiry and Hossain (2019) achieving a 1.4% accuracy in estimating High Aswan Dam's storage and outflow discharge.

For obtaining the water volume of a dam, the level/volume and area/volume curves of the reservoirs needs to be defined which is currently done at the HUS. As mentioned earlier, several studies have been done on the derivation of operation rules of reservoir dams based on modelled inflow combined with satellite data. Although nearly all those papers have a time resolution of 1 month , do have a low resolution and do have lots of data over the dams, making a knowledge gap at the moment when there is nearly no data at all available. Next to this, less papers are published on these topics which contains 10 days average values as desired by the Vietnam Government. (2019)

1.2.3 Knowledge gap:

To my knowledge, little research has been conducted on how operation rules of a reservoir dam in ungauged locations can be derived without the use of in-situ data on a 10-days average scale. This creates a research gap in defining operational rules based on satellite data combined with an assessment of its accuracy without relying on in-situ information for its definition. So, this thesis will be focussing on defining the operation rules of a reservoir dam – located in the Red River system – based on satellite and open source data.

2. Research objective, questions and scope

This chapter gives insight in the research objective and the belonging research questions of this thesis. Starting with the objective, followed up by the questions which supports the objective and the scope and boundaries of this study.

2.1 Research objective

Regarding the previous discussed background together with the knowledge gap, the following research objective is defined: **"The research objective is to set-up and validate a method to derive the operation rules of a reservoir dam by making use of a reservoir dam model combined with remote sensing data for the *Ban Chat* dam in Vietnam and in which in-situ data is used as validation of this rules and method."**

2.2 Research questions

To enhance the understanding of the behaviour of small reservoirs while in-situ data is not available, this study aims to address the following research questions. First datasets about water volumes (WV) over time needs to be formed. Remote sensing data contain some errors and gaps. This makes it important to expand and fill the datasets with different techniques and different sources, resulting in the next question:

- What is the accuracy of a timeseries containing the volume of the reservoir over the time period from construction till now with a resolution of 10 days average, derived from open source satellite data?

The in-situ observed inflow combined with the WV dataset should result in a model in which the operation rules can be derived. The second research question will then be:

- What operations rules simulate the downstream flow of the reservoir dam the most accurate compared with the actual outflow?

At last there remains a question on the accuracy of the model which can give insight in the applicability of the model to different locations.

- What is the accuracy of simulated outflow based on remote sensing data compared with the validation based on in-situ data?

Answering these (sub)questions needs a comprehensive framework. Firstly, the Google Earth Engine (GEE) is used as open source tool to process open source satellite data from the Sentinel-1 and sentinel-2 satellites. Those data sources are combined and gaps are filled to obtain a 10-days average dataset of the water volume of the reservoir. This dataset can then be used together with the in-situ observed inflow data to obtain the historical outflow dataset. This dataset is needed for calibration and validation of the reservoir model in which the operation rules are calibrated and derived.

2.3 Scope and boundaries

This research aims to determine the operation rules of a reservoir dam without relying on in-situ data, enabling potential application to ungauged reservoir dams, by the use of a case study. The methodology integrates inflow observations with reservoir volumes derived from satellite imagery analysis using the Google Earth Engine (GEE) as explained in “Cloud-Based Remote Sensing With Google Earth Engine” (Cardille et al., 2024). Subsequent steps involve testing and refining various operation rules to simulate outflow. The primary objective is to predict reservoir volume fluctuations over a 10-day period, resulting in the calculation of a 10-day average outflow from the reservoir. This time period is chosen based on the National regulations for the *Red River* in Vietnam decided by the Prime Minister of Vietnam (2019).

However, this study is bounded by established algorithms for widely used Sentinel satellites, processed with GEE, for simplicity. Limitations include the intermittent presence of clouds and unsuitable data, requiring a 10-day average model resolution. Furthermore, factors such as evaporation and leakage losses in the reservoir dam are simplified resulting in some small system errors in the output. While this study focuses on evaluating the model's effectiveness for reservoir dams due to the lack of in-situ data, it lays the foundation for future research exploring the model framework's applicability beyond the immediate study area, including potential application to dams outside Vietnam.

3. Data & Materials

This section contains information about the materials used to answer the research questions stated above. Starting with the study area followed by more details about the specific case study locations. Afterwards the needed information about the satellite data is given and this section is ended up by the model information.

3.1 Study area

The study location for this thesis focuses on deriving the operation rules of a reservoir dam situated within a tributary of the *Red River*, the second largest river in Vietnam, which comprises two major tributaries. These tributaries include the *Black River (Da River*) in the south and the *Lo River* in the north, merging to form the *Red River* near Hanoi. Since 2000, new dams have been built in the Chinese and Vietnamese section of the *Red River* with different purposes, being hydropower, irrigation, flood control or multiple purpose dams (Le et al., 2020). All having their impact on the flow regime of the river. Specifically, the study area encompasses the *Da River* basin, where the *Nam Mu River* serves as a national tributary providing accessible in-situ data. Within this basin, the Ban Chat dam is selected for investigation. “The dam supply as peaking plant operating without influences of the complicated water resource management of the international river.” (JICA study team, n.d., pp. 5–2). In other words, the dam is not influenced by upstream water management policies and electricity is only generated when demands are high and therefore ideal as case study to develop a model testing the operation rules based on satellite data.

3.1.1 Reservoir dam: Ban Chat

In this thesis, one dam within this system namely the “*Ban chat*” is chosen as study area. Reason to do this is the availability of the data around this dam. For this location, in-situ data is available in the form of inflow discharge, outflow discharge as well as the water level of the reservoir. This data can be used to validate the model and to check its accuracy.

The *Ban Chat* reservoir dam, is a prominent hydroelectric dam located in a tributary called *Nam mu* of the *Da River* basin able to generate 220MW of power "Ban Chat Hydropower Plant" (2021). Situated approximately 80 kilometres northwest of Hanoi, the *Ban Chat* dam is nestled within the province of Hoa Binh. Besides being an hydro power dam of 132 meter high, the reservoir containing 2.137 km^3 (Prime Minister, 2019) of water serves as a critical water resource for the region supporting agricultural activities, industrial development and urban settlements.

The dam is constructed from 2006 till 2013 (Associates & Associates (MD&A), n.d.) which also have had a big impact on the environment of fish migration, sediment transport and so on, although this will not be mentioned in this report. From 2016, the dam was fully operational.

Figure 3-1: Location of the Ban Chat Dam.
Obtained from Rousseau et al. (2017).

3.2 Data sets

The datasets utilized in this research are detailed below. While this study prioritized the use of open-source data, the in-situ datasets were not publicly available. See table 3.1 for an overview.

Table 3.1: Data sources

Dataset	Source	Retrieval Date	Notes
Sentinel 1 and 2 images	Google Earth Engine. (n.d.). Google. https://earthengine.google.com/	July 04, 2024	Provided by European Union/ESA/Copernicus
Digital Elevation Model (DEM)	Google Earth Engine. (n.d.). Google. https://earthengine.google.com/	July 04, 2024	Provided by NASA JPL
In-situ datasets	Not applicable	Not applicable	Available only to authorized personnel; share on request

4. Methods

First, the key performance indicators (KPI´s) which are used to analyze the timeseries are showed and explained. It then discusses how the Sentinel-1 and Sentinel-2 datasets are used to obtain the water volumes time series with an 10-days average time scale. The validation of this developed fusion method is also explained here. Next, the method for modelling the dam by use of inflow and storage datasets is explained. In this section two models are discussed the water balance model is used to analyse the historical data and to obtain the outflow of the dam while these datasets coming from the water balance are used for the set up of a fuzzy logic model to make prediction for the future based on the operation rules. The last section describes how the final research question can be answered using the observed data from the case study site in Vietnam to validate the generated outflow and with that the operation rules of the dam.

4.1 Key Performance Indicators

This research uses five key performance indicators (KPI´s) while analyzing data, the Nash-sufficient Efficiency (NSE) and the Squared error (R^2) to give insight in the accuracy of mapping the pattern of the time series while the Root Mean Square Error (RMSE), model bias (PBIAS) and Mean Average Error (MAE) gives insight in the correctness of the variability of the model. Table 4-1 gives an overview of the goal, properties and the formula of all indicators. In this table the $Q_{obs/sim}$ are used as example. In this research this variable will be changed in the timeseries on which the formula is applied.

Table 4-1: : Key Performance indicators with their objectives and formulas

	Goal	Properties	Formula by Fernandes et al. (2019)
$\boldsymbol{NSE}$	As close as possible to one	Indicates the quality of the model based on the average of all measured values	$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(Q_{obs}-Q_{sim})^2}{\sum_{i=1}^{n}(Q_{obs}-\overline{Q_{obs}})^2}\right]$
$\boldsymbol{RMSE}$	As close as possible to zero	Indicates the correctness of the model (standard error)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|Q_{obs}-Q_{sim}\right\|^2}$
$\boldsymbol{PBIAS}$	As close as possible to zero	Indicates if the model is likely to over or under estimate the actual values.	$PBIAS = 100x\left[\frac{\sum_{i=1}^{n}(Q_{obs}-Q_{sim})}{\sum_{i=1}^{n}(Q_{obs})}\right]$
$\boldsymbol{R^2}$	As close as possible to one	The R^2 records as a ratio the proportion of the total statistical variance in the observed dataset that can be explained by the model.	$R^2 = \left[\frac{\sum_{i=1}^{n}(Q_{obs}-\overline{Q_{obs}})(Q_{obs}-\overline{Q_{sim}})}{\sum_{i=1}^{n}(Q_{obs}-Q_{sim})^2\sum_{i=1}^{n}(Qsim-\overline{Qsim})^2}\right]^2$
MAE	As close as possible to zero	Gives indication on the average error of the simulation regardless of the direction	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left\|Q_{\mathrm{obs},i}-Q_{\mathrm{sim},i}\right\|$

4.2 Satellite data collection

Determining the timeseries of the reservoir volume over the operation time of the dam is done by the use of remote sensing data. The primary advantage of this approach is its accessibility, providing information on reservoirs without the need for additional local data, which makes it an political independent data source. Various types of satellites can be used to calculated the water surface area. Examples are the satellites from the Landsat and Sentinel program. . The use of satellite is based on the timespan of the research. (baselinegis, 2021) gives an overview of the availability of satellite over time. From this can be concluded that the only usable satellites are the Landsat-7 (L7), Landsat-8 (L8), Sentinel-1 and Sentinel-2. The satellites have several similarities and some differences. (Gerardo & de Lima, 2023) states that Landsat-9 and Sentinel-2 are comparable in the results of water surface mapping. Landsat-9 is not included in this research since the launched in 2021 which gives a too short timeseries for this research. Table 4-2 gives an overview of the most important specifications and (dis)-advantages of the satellites.

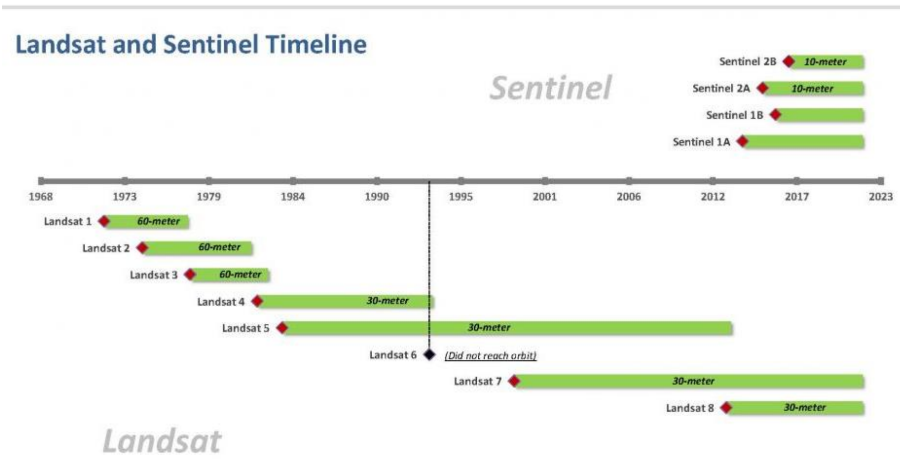


Figure 4-1: Timeline of the Sentinel and Landsat program obtained from Admin1135 (2021).

Table 4-2: Technical information of different satellites.

Satellite name	Spatial resolution	Revisit time	Advantage	Disadvantages
- **Landsat 7**	Optical: 30m	16 days	Multispectral imagery for land and water classification.	Susceptible to cloud cover.
- **Landsat 8**	Optical: 30m	16 days	Improved resolution and sensitivity compared to Landsat 7	Susceptible to cloud cover.
- **Sentinel 1B**	SAR: 5x20m	6-12 days	Penetrating through clouds, rain and darkness.	Low spatial resolution.
- **Sentinel 2B**	Optical: 10-60m	5 days	High resolution	Susceptible to cloud cover.
(Landsat 7 (L7) Data Users Handbook n.d., 8; Landsat 8 (L8) Data Users Handbook n.d., 8; Sentinel-2 products Specification Document 2012; Mondejar and Tongco 2019; Sathianarayanan et al. 2023))				

This study utilizes a diverse array of satellite sensors and platforms, each providing unique capabilities for mapping and analysing water bodies. Landsat-8 OLI (Operational Land Imager) imager, provides high-resolution spatial data at 30m resolution. According to Sathianarayanan et al. (2023), imagery acquired from Landsat-8, which was launched in 2013, offers a long-term time series of earth observation data, facilitating regional research applications.

Landsat-7, equipped with its Enhanced Thematic Mapper Plus (ETM+) sensor, offers multispectral imaging at a 30-meter spatial resolution. Despite technical issues affecting data quality, Landsat-7 data remains valuable for water surface mapping due to its extensive archive and global coverage.

In addition to Landsat-8 data, this study employs Sentinel-2 MSI (Multispectral Instrument) imagery, which offers a comprehensive suite of 13 spectral bands with varying spatial resolutions. These granules cover a ground area of 100 km by 100 km and can be acquired from the GEE. Preprocessing of Sentinel-2 data is required and involved atmospheric correction, image registration, and fusion techniques stated by Sathianarayanan et al. (2023).

Furthermore, Sentinel-1 SAR (Synthetic Aperture Radar) data, obtained from the Alaska Satellite Facility, offering all-weather sensing capabilities for water body extraction as stated by Shen & Fu, (2020).

Based on their respective advantages and disadvantages, Sentinel-1 and Sentinel-2 were chosen as data sources for this research. Sentinel-2 got the highest spatial resolution and Sentinel-1 has the highest temporal resolution. As stated by Du et al. (2016), the Sentinel-2 images would surely be of great significance for regional water bodies' mapping, due to its appealing properties (i.e., the 10-m spatial resolution for four bands and the 10-day revisit frequency) and the free access. Although the issue of cloud cover is still there, especially in a tropical region when cloud cover is high in rain season. This issue is tackled by the selection of the Sentinel-1 data which is unsusceptible to cloud cover.

This research combines the complementary strengths (see table 3.1) of Sentinel-1 and Sentinel-2 to create a highly accurate dataset using a simple method. Sentinel-1 is used as the foundation for the time series due to its short revisit time, which ensures the availability of a large amount of data. In addition, the high spatial resolution of Sentinel-2 is employed to correct the dataset obtained from Sentinel-1. Both satellites requires some preprocessing steps before they can be used and requires several steps. An overview of these steps is shown in Figure 4-2.

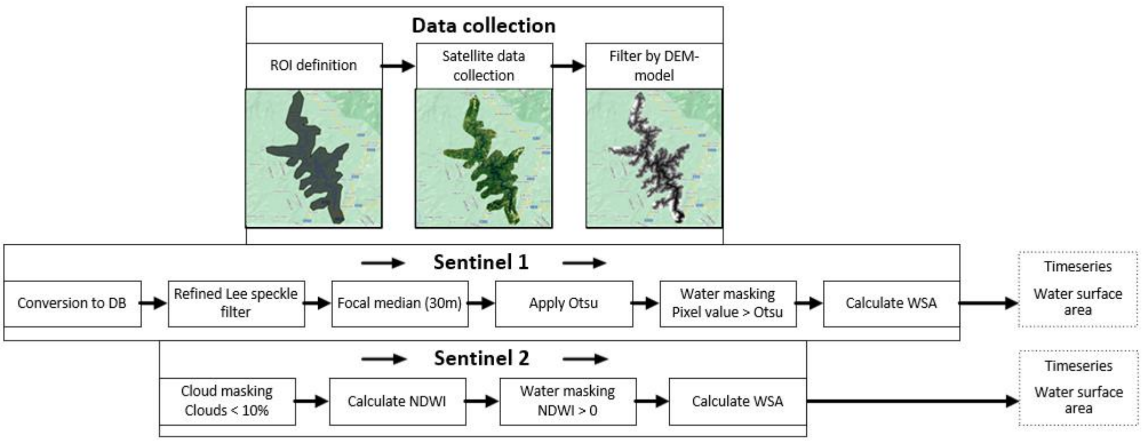


Figure 4-2: Modelling framework WSA extraction

Initially, the region of interest (ROI) is defined with a wide margin around the reservoir to ensure comprehensive inclusion (Step 1) in which the satellite images are being analysed. Subsequently, all available Sentinel satellite images for the ROI between 2016 and 2024 are acquired(Step 2) as well as the Digital Elevation Model (DEM) generated by the Shuttle Radar Topography Mission (Farr et al., 2007) and provided by NASA. This DEM is used to reduce the ROI by filtering out all pixels above the absolute maximum water level of the dam, making the code running faster(step 3). From step 4 Sentienl-1 and Sentinel-2 have different processing steps which are described below, starting with Sentinel-1.

4.2.1 Sentinel-1 Synthetic Aperture Radar (SAR)

Sentinel-1 SAR images are essential for measuring water surfaces due to their all-weather and day-and-night capabilities. These images are captured at different angles during ascending (northward) and descending (southward) satellite passes, affecting the backscatter signal and influencing data interpretation. The angle of incidence impacts how radar signals reflect off water surfaces, crucial for accurate water surface measurements. Therefore, it is important to treat the different paths separately which is done by separating them for the whole process.

SAR images utilize different polarizations, primarily Vertical-Vertical (VV) and Vertical-Horizontal (VH). VV polarization, which transmits and receives signals vertically, is sensitive to surface structures and can introduce noise from vegetation and man-made features. Conversely, VH polarization, transmitting vertically and receiving horizontally, is less affected by structural noise and provides better contrast between water and non-water surfaces. VH polarization effectively distinguishes water surfaces as it minimizes interference from vegetation and other land features. So images are filtered by polarization as well as they path.

Before the Sentinel-1 SAR radar images can be used, they need to be converted to decibels(dB) to get a better differentiation between surfaces. Afterwards they are subjected to filtering processes to eliminate noise. This can be done by different methods. In this research, the best results are obtained by using the refined Lee filter algorithm applied in the GEE, as specified and explained by Qiu et al., (2004). After applying the refined Lee filter, a focal median filter is used to eliminate any remaining noise pixels. This filter considers a neighbourhood around each pixel (30 m) and replacing the pixel value with the median of the values in this neighbourhood (Qiu et al., 2004). These steps result in a clean image collection of which the water surface area can be extracted based on a threshold. This research utilizes dynamic Otsu thresholding, since it does not requires prior knowledge about the image and it can determine the optimal threshold value automatically (Helmy, 2023).

4.2.1.1 Otsu thresholding

In Step 7, the Otsu thresholding algorithm (Otsu, 1979) is applied to determine the optimal threshold for distinguishing water from land. The Otsu thresholding method is widely used in surface water mapping with Sentinel-1 SAR images (Tan 2023). It has been particularly effective in reservoir areas, with a dynamic Otsu thresholding algorithm showing high accuracy in surface water mapping and flood monitoring (Tran 2022). This method is particularly advantageous in remote sensing and SAR data processing because it does not require prior knowledge about the image's histogram (Figure 4-3). By effectively distinguishing between water and land pixels, Otsu thresholding enhances the accuracy of Water Surface Mapping (WSM), making it a robust choice for diverse and complex datasets. An optimal threshold value for every image is defined which is used to classify pixels into water or non-water pixels.

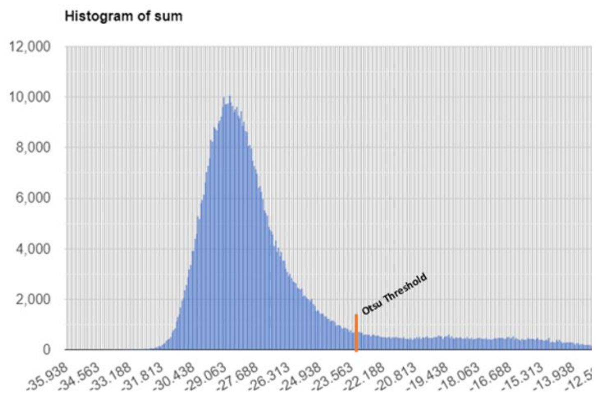


Figure 4-3: Example of the image histogram with an Otsu adaptive threshold in it

Finally, all pixel values below the threshold are classified as water and are aggregated and multiplied by the pixel area to compute the total WSA. This procedure is applied to all images from 2016 to the present, and separately for ascending and descending paths. Both paths are sufficient for determining the WSA, although there may be some discrepancies between them, as noted by Mateus et al. (2017). Resulting in two distinct datasets, each containing the date of the image along with the corresponding WSA.

4.2.2 Sentinel-2 imagery

Sentinel-2 imagery requires some other steps to define the WSA. These images exists out of visual bands, so clouds can have a significant influence on processing steps making Images with more than 10% cloud coverage being excluded from the data collection (step 4).

The resulting dataset is analysed using the normalized difference water index (NDWI) as proposed by (McFeeters, 1996)**.** This index is designed to maximize the reflectance of the water body in the green band as well as to minimize the reflectance of water body in the Near Infra-Red (NIR) band and is calculated as:

$$NDWI = \frac{X_{green} - X_{nir}}{X_{green} + X_{nir}} \qquad \text{Equation 1}$$

The X_{green} and X_{nir} are the reflectance values in the green and near infrared wavelengths, respectively. And are corresponding to Band 3 and Band 8 from the Sentinel-2, both having a spatial resolution of 10 m and so the calculated NDWI in equation 1 has a spatial resolution of 10 m. After the NWDI imagery for al images is produced, water maps are made. A threshold of 0, proposed by McFeeters, is used to distinguish between water and non-water pixels. Pixel containing values above 0 are classified as water and less than 0 are classified as non-water. The number of water classified pixels multiplied by the surface area of a pixel (100 m^2) is then the estimation of the WSA.

4.2.3 Exporting and building 10-days average timeseries

When both data sources are analysed, datasets containing water surface areas on different days are obtained. This study aims to create a comprehensive WSA dataset with a 10-day temporal resolution, necessitating the transformation of the dataset to this timescale. We established a list of start dates for 10-day periods, beginning each January 1th each year. This uniform temporal framework facilitates straightforward comparisons across different datasets and time periods. For each satellite source, we verify if a WSA is available within the predefined 10-day periods. When multiple WSA values were available for the same 10-day period from a single satellite, we calculated the average, to ensure temporal alignment. This step ensures that all data points are aligned temporally. The final output consists of three datasets, each containing 10-day average WSA measurements from different satellite sources. These datasets are harmonized for consistent temporal resolution, enabling direct comparison and comprehensive analysis. The three datasets are utilized:

- Ascending Sentinel-1 SAR data:
  This dataset is chosen as the base due to its extensive coverage, containing data for 69.8% of the time periods from January 1, 2016, to March 19, 2024.
- Descending Sentinel-1 SAR data:
  This dataset has a coverage of 50.2% over the same period.
- Sentinel-2 optical imagery:
  Complementary data source, not primarily used for filling gaps but for calibration and additional analysis.

Both the ascending and descending datasets have gaps every six rows due to the 12-day revisit time of the Sentinel-1 satellite. The gaps need to be filled to ensure a continuous 10-day interval dataset. The ascending dataset is selected as the base because it has the highest coverage (69.8%). The descending dataset often has data available when the ascending dataset does not, due to the offset between their paths.

Filling up the gaps in the ascending dataset is done by using the descending value for that time period corrected by the average difference between ascending and Descending values. This process increases the coverage of the ascending dataset to 87.7%. The last gaps are filled by linear interpolation.

4.2.4 Data fusion.

The results of Water Surface Mapping (WSM) in km^2 from Sentinel-2 are generally more accurate than those from Sentinel-1. Souza et al., (2022) and Peña-Luque et al. (2021) also indicates that Sentinel-1 likely underestimates the surface area) This research uses the Sentinel-2 data as correction of the Sentinel-1 data. Figure 4-4 shows a correlation of 0.89 between the Sentinel-1 and 2 observations making it possible to correct the Sentinel-1 values based on the Sentinel-2 observations.

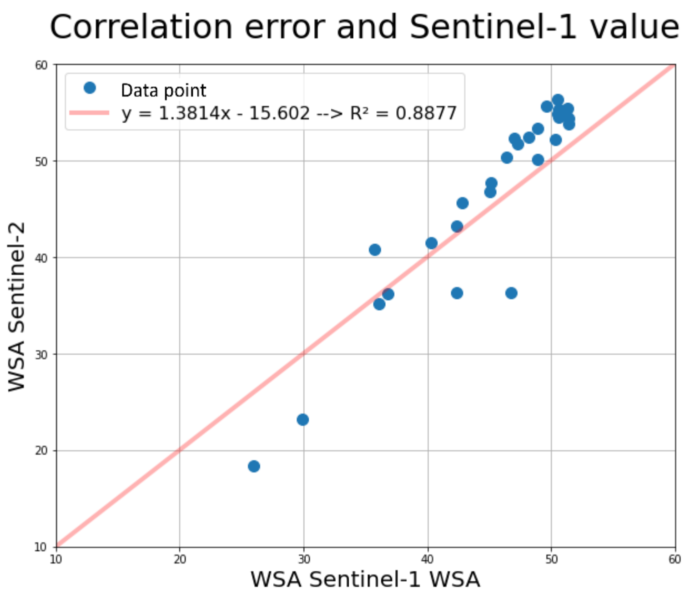


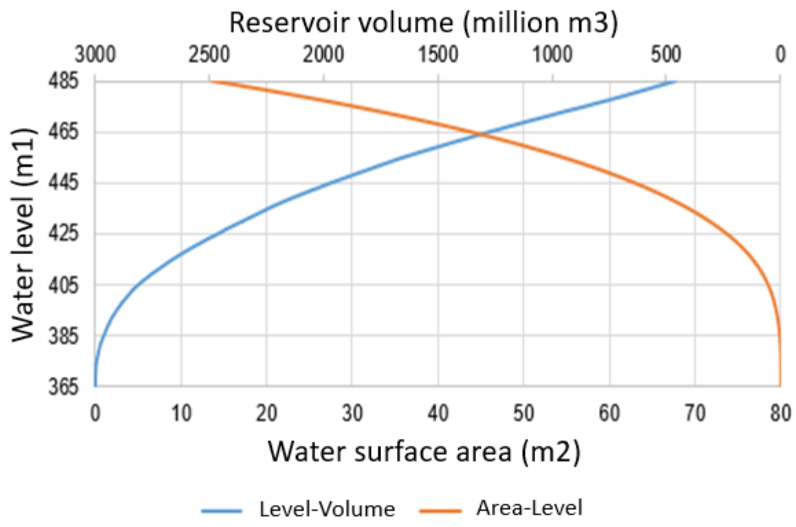
Figure 4-4: Fusion method of the data from 2 satellites, based on the correlation.

To do so, the linear relation between the two sources is used as correction formula, shown below. This results in a dataset of water surface areas obtained by Sentinel-1 and corrected to Sentinel-2.

$$WSA_{corrected} = (1.568 \times WSA_{sentinel1} - 24.73) \quad in\ km^2 \qquad \text{Equation 2}$$

4.2.5 Water Volume extraction

The ultimate objective of delineating the surface area of reservoirs lies in deriving their corresponding volumes. This process is facilitated through the utilization of water-area-level-volume curves, which have been established by Mr. Minh (2024, in preparation) based on digital elevation models (DEMs). In order to obtain the volumes, the surface area must first be converted in to water level, which can subsequently be transformed in to a volume. Mr. Minh (2024, in preparation), provided formulas which can do it automatically, shown in equation 3 and 4. This will give one dataset, containing the water volume with a temporal scale of 10 days.



Bản Chất reservoir

Figure 4-5: Water Area-Level-Volume curves, made by Mr. Minh (2024) and data obtained from the decision statement of the Prime Minister of Vietnam (2019).

$$Y = -1 + 2 \times (WSA - \frac{0.003}{67.846 - 0.003}) \qquad \text{Equation 3}$$

$$WV = 129.63 \times Y^4 - 20.44 \times Y^3 + 241.16 \times Y^2 + 1259 \times Y^1 + 868.77 \qquad \text{Equation 4}$$

The WSA and the WV dataset are validated by the use of the KPI's explained in section 4.1. The in-situ observed data uses the same algorithm (section 4.2.3) to make a 10-day temporal scale timeseries of it is used as observed dataset.

4.3 Reservoir model

This section shows how the reservoir model is buil1, which variables there are and how they are calibrated. First the timeseries obtained by previous explained methods are used to construct a water balance which is used to analyse flow and to obtain an outflow timeseries which can be used in the Fuzzy model. This fuzzy model is used to derive operation rules. Then the Fuzzy model is explained including several input variables. The final calibrated fuzzy model is able to simulate the discharge downstream of the reservoir dam. The validation of this model is explained in the next section.

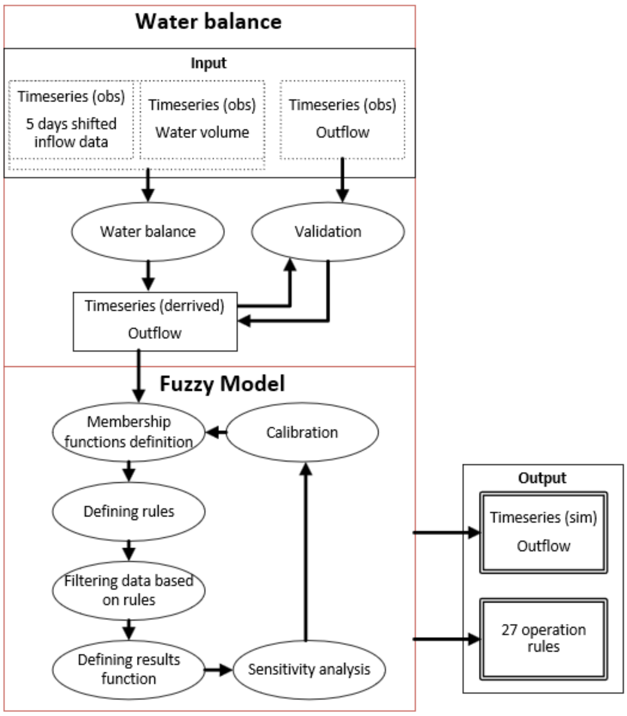


Figure 4-6: Modelling framework reservoir model

4.3.1 Water balance

The water balance is a fundamental element in hydrology used to assess the availability and movement of water within a specific system. This approach involves quantifying the inflow, outflow, and storage changes of water to understand the overall dynamics and sustainability of water resources. The inflow and outflow of this system consists of several flows such as leakage, seepage, rainfall inflow and release discharge in this study only inflow and release discharge are considered, resulting in the water balance equation 5. This balance can be rewritten to equation 6 in which the release is a function of the known inflow and water storage.

$$I(t) \times t + S(t) - R(t) \times t = S(t+1) \qquad \text{Equation 5}$$

$$I = Inflow \left(\frac{m^3}{period}\right)$$
$$S = Storage\ (m^3)$$
$$R = Release \left(\frac{m^3}{period}\right)$$
$$t = Time\ period\ (10\ days)$$

$$R(t) = I(t-1) + S(t-1) - S(t) = I(t-1) + \Delta S \quad \text{Equation 6}$$

The average water volume is defined over 10-day periods starting on the 1st of January every year, and the inflow in the mass balance should represent the average of inflow between these two time periods. And so 10-day periods for inflow and outflow starts at the 6^{th} of Januari. Equation 9 shows how the 10 days average inflow is converted to a total inflow volume per 10 days. After conducting these steps, timeseries of the inflow, water volume and outflow are obtained to build the Fuzzy model on.

4.3.2 Fuzzy logic model

After the historical data is collected using the water balance, fuzzy logic is applied to build a model that can be used to make predictions for the future. Fuzzy logic has been identified as a valuable tool in addressing uncertainty and complexity in environmental and water resource systems modelling (Mujumdar & Ghosh, 2008). This is particularly relevant in the context of water inflow, volume, and outflow data, where the lack of a clear relationship can be addressed through the use of fuzzy logic to incorporate uncertainty and imprecision (McKone, 2005). Fuzzy logic-based models have been successfully applied to simulate water dynamics in soil, demonstrating their potential in addressing similar challenges in water systems (Freire, 2014). Furthermore, the transparency and ability to incorporate qualitative knowledge and uncertainty make fuzzy logic a promising approach for deriving operation rules in river management (Janssen, 2006). And so for this research it is a promising approach for deriving operation rules in reservoir dams.

The key ideas are that fuzzy logic allows for something to be partly this and partly that, rather than having to be either all this or all that; and that the degree of "belongingness" to a set or category can be described numerically by a membership number between 0 and 1. As an example, Fig. 1 shows typical membership functions for low, medium, and high outflow volumes. In this example, demand less than 50 is defined as “low”, between 75 and 125 as “medium” and above 150 as “high”. In fuzzy logic terms, a demand of 60 has a membership of 0.28 in “low” and 0.20 in “medium”. The membership function (MF) is made linear to make subsequent calculation easier.

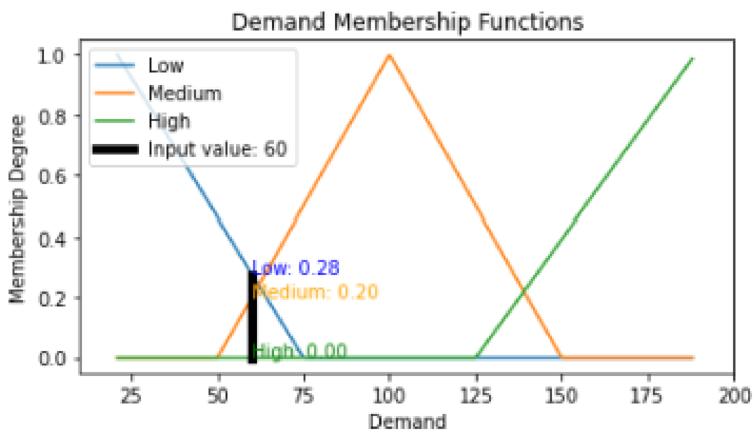


Figure 4-7: Example explanation of the membership functions

To derive the operation rules, fuzzy rules can be established using input and output variables. The objective is to obtain rules that can accurately simulate the outflow, with the outflow set as the output variable. To identify the key influential variables, an extensive data analysis was conducted. This involved plotting pairs and triplets of variables against the corresponding outflow to observe any emerging patterns. If a discernible pattern was found, it indicated that those variables were important for the study. Through this analysis, water inflow, volume, and demand were identified as the primary input variables. Important to note is that the demand is not the actual demand but is assumed to be the long term average for the specific time in the year. This is done as the data analysis shows a quite consistent pattern over the years which can be encountered by a demand curve which is yet unknown. The three input variables, are divided into three categories, and simple triangular MF are used as shown in Figure 4-7: Example explanation of the .

Fuzzy rules are set up in the form: "if the value of variable $Inflow_1$ is "low" and variable $Volume_1$ is "low" and $Demand_1$ is "medium" then the consequence is "medium". As first step, 3 variables are chosen as check for the method. The categories are giving the names "low", "medium" and "high". The number of rules in the control system is $3^3 = 27$ The rule base have the form of:

$$inflow_t\,[\text{"low"}] \;\;\&\;\; volume_t\,[\text{"low"}] \;\;\&\;\; demand_t\,[\text{"medium"}] \;=>\; outflow_t\,[\text{"low"}]$$

The category of the result also needs to be defined which can be done in several ways. This paper utilizes the data driven approach which means that the result of the rule is defined based on the dataset of derived outflows. First a calibration dataset is made. This is done by taking 70% random samples from the overall available data. The reason to choose for random samples is the yearly trend in the data together with the influence of unknown events in the data which can affect the quality of the calibration as well as validations by taking a continuous timeseries. But before splitting up, the moving average of 1 forward and 1 backwards of the inflow and outflow is taken to reduce the noise in the dataset. Only 1 step for and backwards is chosen to contain the small fluctuations and keep the accuracy as high as possible without flattening out to much. The calibration dataset is filtered based on the conditions given in the rule. The average value of the outflow is calculated, the membership to the three categories are defined and the category with the highest membership is used as result of the rule.

If a set of input variables has a membership in multiple categories, multiple rules are triggered. For example the $demand_1$ has a membership of 0.28 in "low" as well as 0.2 in "medium" so 2 rules are triggered based on demand. For every triggered rule, membership values are computed and multiplied with the result of the consequences assigned by the rules. The output of the rules are combined by a defuzzification method to give a final single output. There are multiple defuzzification methods available, the best results are obtained by the centroid method which is shown by equation 8. $w_{t_{i,v,d}}$ is the weight of the rule while $Outflow_{t_{i,v,d}}$ is the output value of that certain rule. So the output of the model on timestep t will be a combination of several weighted outputs.

$$Outflow_t = \frac{\Sigma(w_{t_{i,v,d}} \times Outflow_{t_{i,v,d}})}{\Sigma(w_{t_{i,v,d}})} \qquad \text{Equation 8}$$

4.3.3 Global Sensitivity analysis of fuzzy model

Before calibrating the model, the calibration parameters should be defined. The rules are defined in an automated, data-based manner using the membership functions (MF), making the MF the sole calibration parameters of the model. Due to time constraints, only MF comprising of three triangulars are analysed. For all parameters, the peak of the three categories together with the width of the triangulars are used for the sensitivity analysis. The parameters for the MF of inflow(i), volume(v), demand(d) or outflow(o) combined with the category low(l), medium(m) and high(h) and a number in which 1 stands for left corner, 2 for the top and 3 for the right corner. So the left corner of outflow high will be oh1 (Outflow High 1). “ow” stands for outflow width.

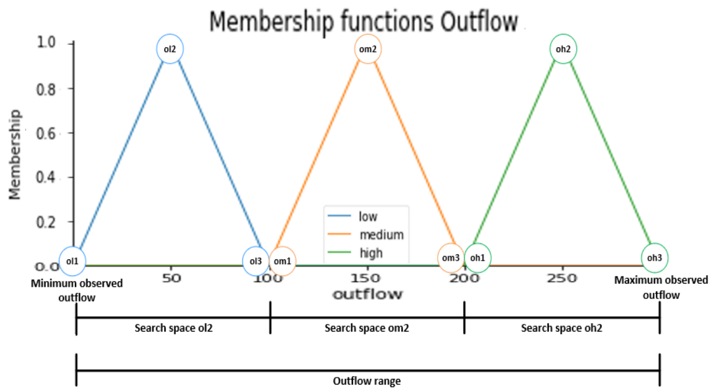


Figure 4-8: Coordinate definition for the membership functions

Since the relationship between the variables is unknown but likely significant, a global sensitivity algorithm that considers parameter interaction is required. The Morris method is chosen to apply. It is a fast and reliable sensitivity algorithm which takes the interaction as well as the nonlinearity into account according to Song et al.(2015). This analysis requires the parameters to have a certain range in which they can move. The ranges of the parameters are shown in table 4-3. The outflow variable is taken as example, other variables ranges are defined exactly the same. The method requires a result value on which the sensitivity needs to be measured. This research is focused on obtaining the operation rules of the reservoir. Since the NSE PKI is the best indicator for the correctness of the pattern, this is used for the sensitivity analysis as well as for the optimisation in the next section.

Table 4-3: Boundaries of the membership functions

PARA-METERS	BOUNDARIES	TRIANGULAR MEMBERSHIP FUNCTION VALUES
OL1		0
OL2	[Minimum observed outflow, Minimum observed outflow + 1/3 outflow_range]	$ol2$
OL3		$ol2 + 1/2 * ow$
OM1		$min(ol3, max(0, om2 - 1/2 * ow))$
OM2	[Minimum observed outflow + 1/3 outflow_range, Minimum observed outflow + 2/3 outflow_range]	$om2$
OM3		$om2 + 1/2 * ow$
OH1		$min(om3, oh2 - 1/2 * ow)$
OH2	[Minimum observed outflow + 2/3 outflow_range, Maximum observed outflow]	$oh2$
OH3		$\max(oh2 + 1/2 * ow, statistics\_matrix.loc["Maximum", "outflow"])$
OW	[outflow_range/2, outflow_range]	

Running the Morris analysis for a sample set of 1000 with numerical range of 12 gives insight into the sensitivity of every parameter in terms of σ and μ. Parameters with a low μ are less sensitive then values with a high μ. Parameters with a low σ behave linear while high σ indicates an non-linear behaviour of the parameter. And μ/σ shows the robustness of the Parameters. Parameters with high μ and high μ/σ are taken to be the calibration parameters.

4.3.4 Global optimisation fuzzy model

The calibration of the parameters is done by the Shuffled Complex Evolution optimization algorithm by Duan et al. (1993) and applied in the Spotpy library Wu & Zhu, (2006) shows the relevance of this algorithm. It is an global optimization method that does not rely on explicit expressions for the objective function or its derivatives, making it suitable for handling nonlinear problems with high-parameter dimensionality. This algorithm requires a range for the parameters in which they can be optimized. The range is set to be 50% lower or higher than the initial values, coming from the best performing parameters in the sensitivity analysis. Calibration of the model takes still a lot of time, making the number of runs set to 1000 after which the new parameters are obtained and used to run the calibration algorithm again but with a lower range of 10% on the optimized values.

Table 4-4: Calibration range MF parameters

PARAMETER	LOWER BOUNDARY	UPPER BOUNDARY
DM2	0.50 * dm2	1.5 * dm2
DW	0.50 * dw	1.5 * dw
OL2	0.50 * ol2	1.5 * ol2
OM2	0.50 * om2	1.5 * om2
OH2	0.50 * oh2	1.5 * oh2
OW	0.50 * ow	1.5 * ow
IL2	0.50 * il2	1.5 * il2
VW	0.50 * vw	1.5 * vw
DH2	0.50 * dh2	1.5 * dh2

4.4 Reservoir model validation

The model validation process utilizes two distinct time series: outflow and water volume. These datasets are randomly sampled from the available data, following the calibration section's approach where 70% of the data is used for calibration and the remaining 30% for validation. Both validation procedures employ Key Performance Indicators (KPIs) previously described. By comparing the KPIs derived from these two validation methods, insights into the model's behaviour are gained, evaluating whether volume-based KPIs reliably reflect the model's performance. Outflow data KPIs serve as reference benchmarks for this comparative analysis. Figure 4-9 shows an diagram of the applied method.

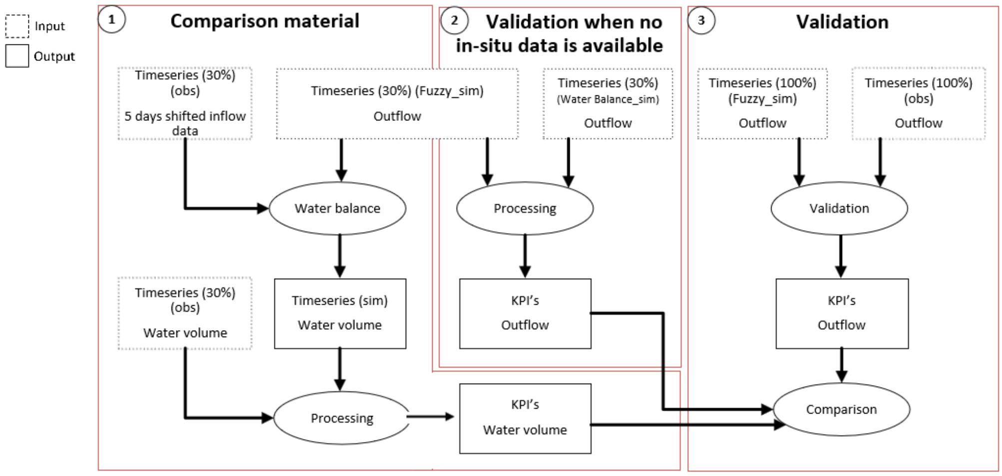


Figure 4-9: Modelling framework for the validation

First, the validation of the fuzzy model is established by using the by fuzzy modelled outflow to compare it with the observed outflow of the whole period. This is visually as well as statistically validated by the use of a graph and the KPI's.

Secondly, the observed inflow data and the simulated outflow data from the Fuzzy model are incorporated into the water balance calculations to generate a simulated time series of water volume. This simulated water volume time series is then normalized and compared with the normalized observed water volume time series to define the KPIs. Normalizing both volume series enhances the comparability of the different validation methods. Similarly, the outflow time series for the validation dataset are compared with the outflow of the water balance simulation, resulting in two sets of five KPIs each. The KPIs based on the water balance outflow and the KPIs from the volume time series are used for comparison with the validation to enhance understanding and insight into the model when in-situ data is not available for validation.

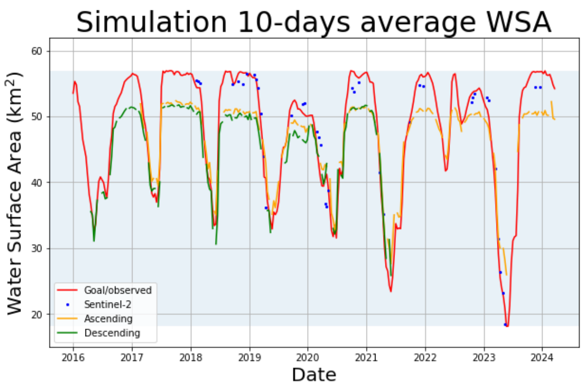
5. Results

The results are discussed in the order of the research questions. So first the results of obtaining a water volume set is shown in section 5.1. The model which is made of the reservoir dam is discussed in section 5.2 and the analysis on the reliability of the model and insight in the behaviour of the model when there is no in-situ data available is presented in section 5.3.

5.1 Reservoir volume datasets

The observed and by satellite calculated WSAs for the entire operational period of the dam are shown in Figure 5-1. It can be concluded that all three methods effectively map the dynamics of the WSA. Peaks are generally timed correctly, although the derived values do not always reach the correct levels. This is supported by the KPI results listed in Table 4-1: : Key Performance indicators with their objectives and formulas. As in the hypothesis in the methods section, Sentinel-2 performs the best in mapping the WSA, which can be concluded visually as well as by the KPI scores. According to the KPI's, the Sentinel-2 is the best in mapping the pattern due to its high NSE-value as well as very accurate in obtaining the variability of the data indicated by R^2.

However, the combination of different sources outperforms each individual source in almost all areas. Only the PBIAS of Sentinel-2 shows a better performance than the Fusion series. However, this indicator is less important due to the interest in the pattern of the time series. Therefore, the most important indicators are the NSE and R2 values. The overall high availability of the Sentinel-1 data is combined and corrected to the performance of the Sentinel-2 measurements. Resulting in an even better performing dataset based on the original dataset, which is in line with the hypothesis from the methods section. It can even be concluded that the combination of both satellites outperforms the single use of them, making this method very suitable for measuring the WSA with a 10-day temporal resolution in ungauged locations.



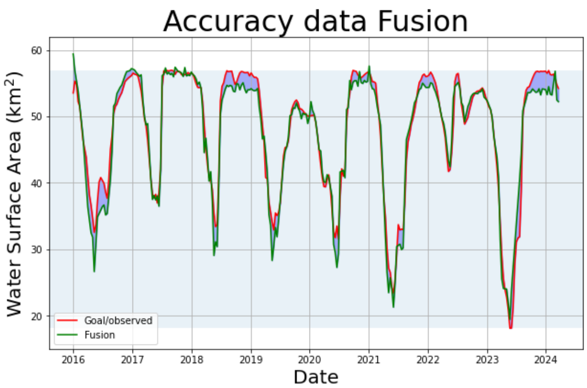


Figure 5-1: WSA for all sources (left), WSA for the fusion set (right)

Table 5-1: table shows the results of the KPI's for all data sources in calculating the WSA over time compared with the observed dataset.

Ban Chat reservoir (unit)	NS (-)	PBIAS (-)	RMSE (km^2)	R2 (-)	MAE (km^2)	Mean (km^2)	Median (km^2)
Area-Sentinel_2	0.89	0.21	3.47	0.94	2.43	47.11	52.22
Area-Descending_server	0.77	6.48	4.05	0.98	3.52	45.51	47.83
Area-Ascending_server	0.83	3.82	3.80	0.99	3.30	46.39	49.06
Fusion	**0.96**	**-0.61**	**1.92**	**0.98**	**1.47**	**48.23**	**52.12**
Observed	-	-	-	-		51.47	47.89

5.1.1 Water Volume

The water volume time series is presented in Figure 5-2. It closely mirrors the data fusion time series of the WSA, which is expected since a single formula is used to convert the WSA into water volume. The method tends to overestimate the water volume in the reservoir, both at high and low water levels. There are two general explanations for this behaviour:

The overestimation at low water levels might stem from the computation of the observed data. The observed water volume is derived by measuring the water level in the reservoir and converting it into water volume using Figure 4-5. The error margins as well as the exact location of these measurements are unknown, making it challenging to draw final conclusions about the differences.

Another explanation for the discrepancy lies in the origins of the measurements. The observed data measures the water level, while the satellites measure the WSA. Satellites may detect more water in the study area due to the presence of separate pools, especially during periods of low water levels. Resulting in a higher WSA as well as higher water volume. Obtaining more insight in this can be obtained by the use of different satellites and is further discussed in the recommendations section.

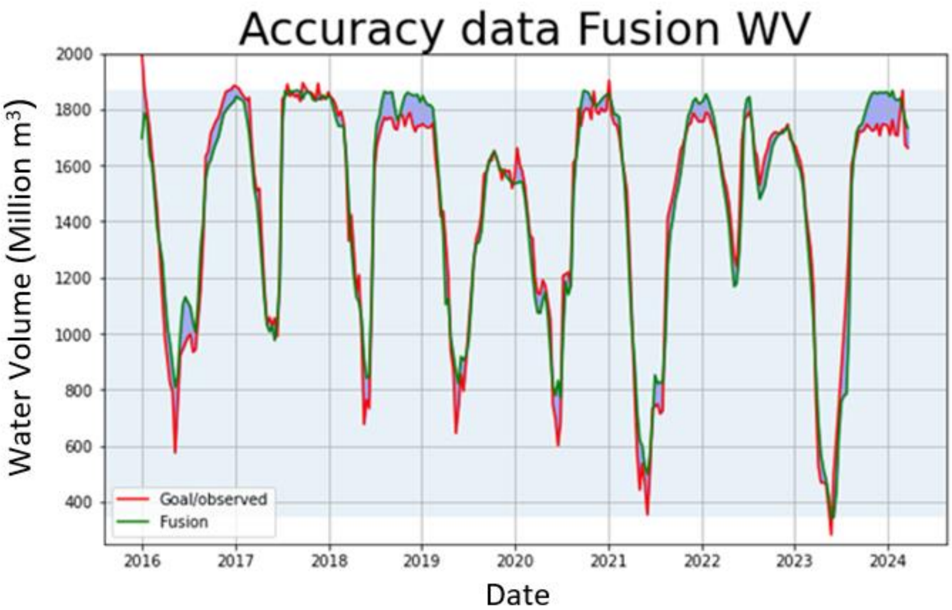


Figure 5-2: This graph shows the calculated and observed timeseries of the water volume in the Ban Chat reservoir.

Despite the tendency to overestimate the water volume compared to the in-situ measurements, it can be concluded that the used approach of calculating the water volume of the Ban Chat reservoir performs very good in mapping the pattern of the storage while it has some small limitations in mapping the height of the peaks, with an overall mean average relative error of only 4.50%.

Table 5-2: Water volume validation results in KPI's

Ban Chat reservoir (unit)	*NS* (-)	*PBIAS* (-)	*RMSE* (10^6 m^2)	*R2* (-)	*MAE* (10^6 m^2)	*Mean* (10^6 m^2)	*Median* (10^6 m^2)
Water Volume	0.96	0.27	84.05	0.98	66.15	1,468.10	1,623.53

5.2 Reservoir model operation rules

This section discusses the model of the reservoir dam. This is done in the same order as in the methodology section. First the results of the water balance are shown. The made water balance by the use of observed inflow and by remote sensing obtained WSA is a sufficient way to obtain the outflow for a historical dataset. Proven by a NSE score of 0.83 with a likelihood to overestimate the flow with 1.93%. The calibrated fuzzy logic model has lower scores in simulating the outflow of the reservoir dam, namely 0.59 and a likelihood to underestimate the flow with 4.13. Main findings coming from this research are the applicability and promising fact of the implementation of fuzzy logic in modelling reservoir dams in ungauged locations. Several statistical datasets can give insight in the behaviour of the model when in-situ data is not available to validate the model.

5.2.1 Water balance results

The water balance without considering the precipitation and evaporation shows a average relative error of 1.55% is a sufficient method to obtain water discharges for the downstream of the dam. Figure 5-3 shows the moving average of the observed outflow together with the mean average of the by water balance simulated outflow. There are a few points to discuss. In general the modelled time series is fitting the pattern of the observed data. When outflow is low, the model seems to overestimate the outflow which can be the result the assumption to not encounter the evaporation in this model. Most of the time, when the outflow is low, the water level is high making a big surface area so also more evaporation. Although in the dry periods, when water level is low, but the potential evaporation rate is higher, the model does not seem to have a consistent over or under estimation. Making the evaporation not an important factor to take into account in this model.

In 2020 the model over estimates the flow with more than 100%. A cause of this can be apparently big evaporation although we assumed evaporation to have small influence in this reservoir water balance model. Besides the model, the measurements do also have errors in it. By looking at the pattern of the yearly graph, the difference might be caused by errors in the measurement station, however more research needs to be conducted on this.

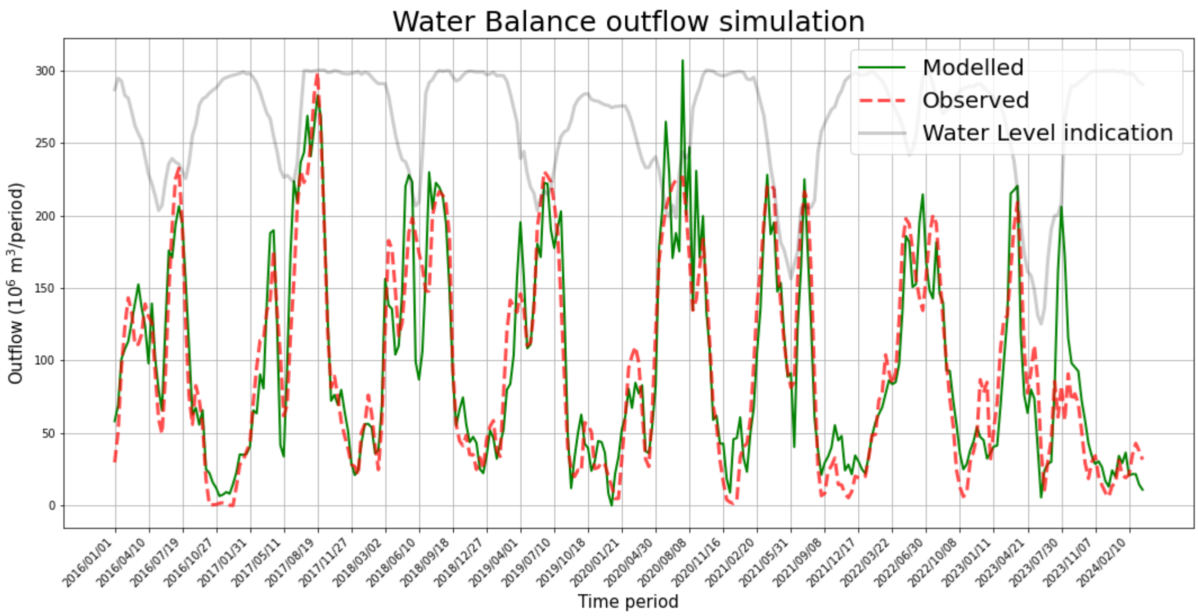


Figure 5-3: Downstream discharge modelled by the water balance

The statistics of the model gives some more insight in the general performance of the model. Showing a high NSE and R2 value indicating the model to be a good representation of the observed values and sufficiency in mapping the pattern of the outflow. With a likelihood to overestimate the outflow by 1.93% with a mean absolute error of 1.52 (10^6 m^3) makes the model reliable enough to use for calibration of the fuzzy model in predicting the outflow of the reservoir dam.

Table 5-3: Results of the simulated discharge validation in KPI's

obs_BC1	***NS*** *(-)*	***PBIAS*** *(-)*	***RMSE*** *(10^6 m^3)*	***R2*** *(-)*	***MAE*** *(10^6 m^3)*	***Mean*** *(10^6 m^3)*	***Median*** *(10^6 m^3)*
MA_WaterBalance_sim_outflow	0.83	-1.93	29.10	0.91	1.52	98.08	76.00

5.2.2 Fuzzy Logic Model Results

The results of the calibration of the MF's are depicted in Figure 4-9, illustrating a scatter plot of the calibration. Besides the NSE-value, the MAE is taken into account by choosing the best fitting set of parameters. The MAE should be as low as possible while the NSE should be as close as possible to 1. There are 2 options making a pareto front in this case. Indicated with green. Since the difference between both is negligibly small, the highest NSE-value is chosen to be the best parameter set.

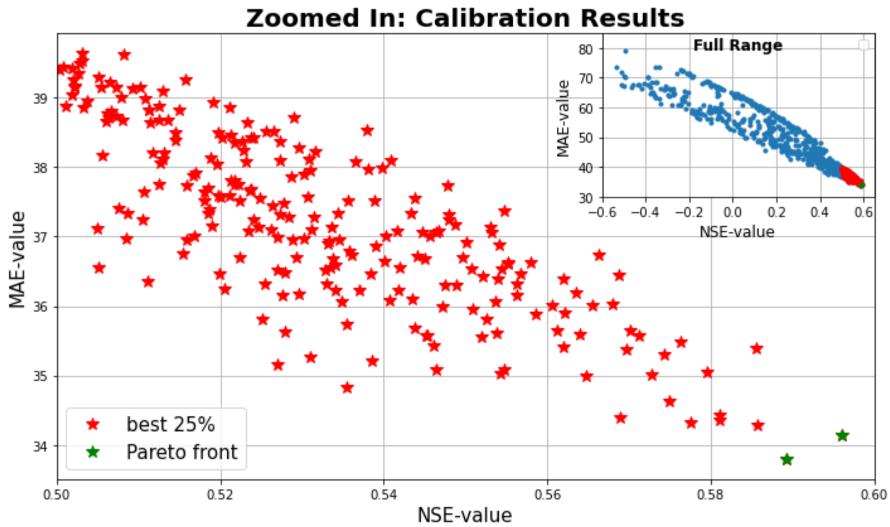


Figure 5-4: Results scatter plot of the calibration

With these results, the fuzzy logic model is a promising method in modelling the reservoir dam outflow in ungauged locations. A simple membership functions set up shows a small increase of the outflow results compared to the average outflow of the period based on the simulated outflow. This is supported by the slight improvement in PKI's shown in Table 5-34. The estimation of the pattern is slightly increased, supported by a better NSE and R2 value while the model has a higher likelihood to underestimate the outflow downstream. Together with the slightly lowered MAE, it can be concluded that the Fuzzy model has potential in modelling the downstream flow of the dam. It is evident that the fuzzy model primarily follows the pattern of the demand curve, which is explainable by the fact that the demand curve is chosen to be the average of outflow per timestep. Notably, the simulations produced by the model exhibit a stepwise appearance. This phenomenon can be attributed to the definition of the membership functions (MF) and their associated boundaries. These outcomes suggest that employing three triangular membership functions per variable fails to sufficiently differentiate between the various situations that arise. To achieve better results, alternative membership functions such as trapezoidal shapes, as well as an increased number of membership functions, should be considered. This is further elaborated in the recommendations section.

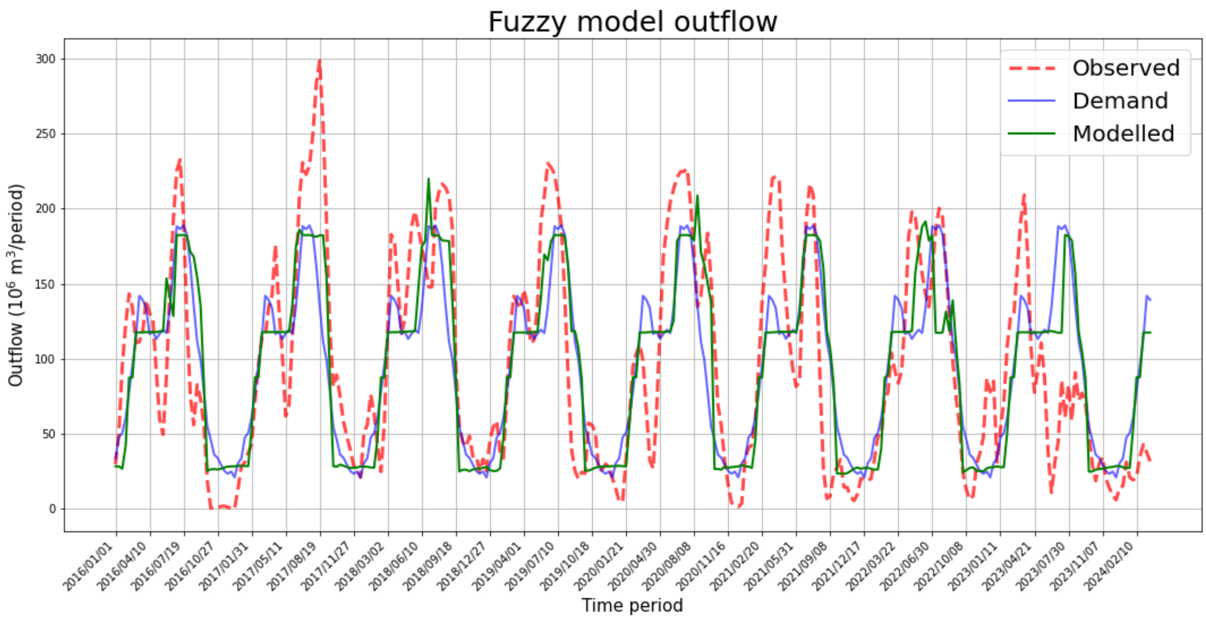


Figure 5-5: Fuzzy model simulated outflow

Table 5-4: Accuracy of the simulation of the fuzzy logic model

obs_BC1	***NS***	***PBIAS***	***RMSE***	***R2***	***MAE***
Fuzzy model predictions to water balance	0.59	6.26	46.48	0.78	34.47
Demand to water balance	0.55	3.40	48.90	0.74	34.99

The operation rules supporting this fuzzy model are shown in Table 5-4. From the sensitivity analysis can be seen that the outflow and the demand are the most important variables for the model. Which can also be concluded from the rules. When the demand is high, the outflow is probably also high, while the inflow and the volume does not seems to have a big influence on the outcome of the rules. The result of the general rules is in line with the hypothesis. The outflow tends to follow the demand curve while the outflow does not rely on the demand anymore when the volume as well as inflow are high since then as much water as possible should be released to keep the water level below the emergency level.

Tabel 5-5: Derived operation rules in fuzzy logic format. Colours are for clarification.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
infl	l	l	l	l	l	l	l	l	l	m	m	m	m	m	m	m	m	m	h	h	h	h	h	h	h	h	h
volume	l	l	l	m	m	m	h	h	h	l	l	l	m	m	m	h	h	h	l	l	l	m	m	m	h	h	h
demand	l	m	h	l	m	h	l	m	h	l	m	h	l	m	h	l	m	h	l	m	h	l	m	h	l	m	h
Outfl	l	m	m	l	m	m	l	m	m	l	m	m	m	m	m	h	h	h	l	m	h	m	m	h	h	h	h

This membership functions used to define the rules are shown in figure 5-6. The horizontal axis shows the input value of the variable while the vertical axis defines the membership in a certain function. The combination results in the addressed rules which is also a membership function for defuzzification shown in figure 5-6 (right bottom). The exact coordinates of the MF, obtained after the sensitivity analysis, table 5-5.

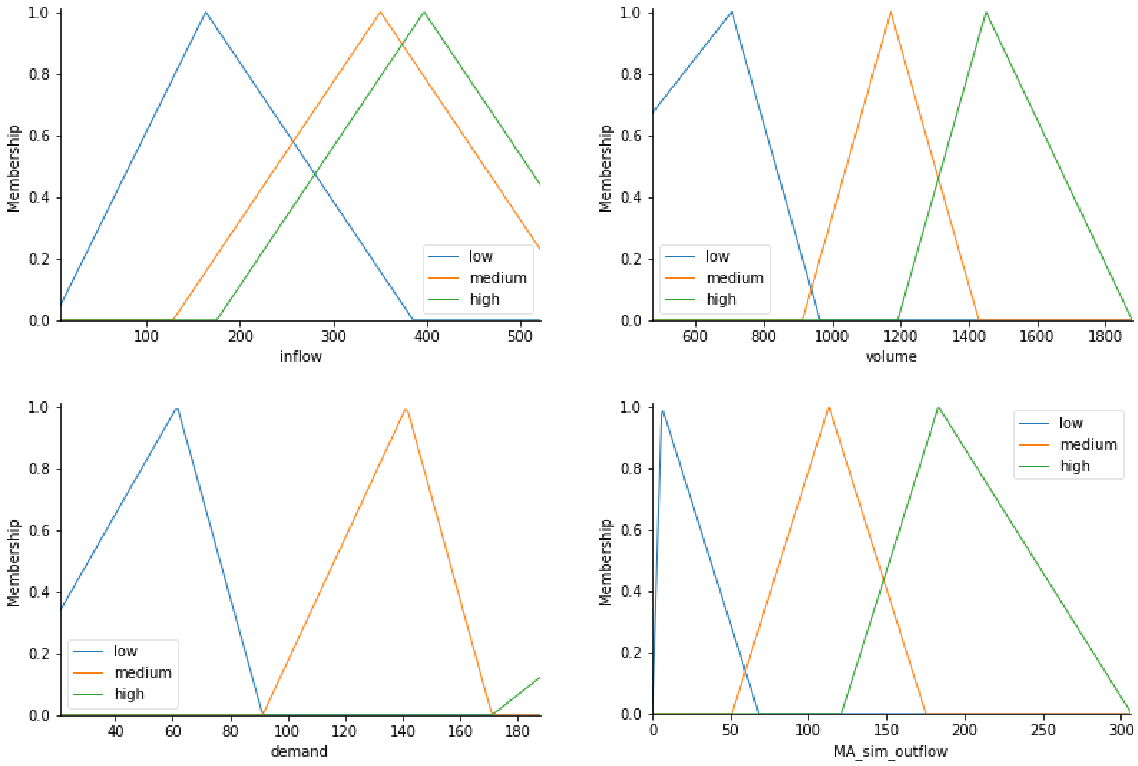


Figure 5-6: Membership Functions displayed, left top: Inflow, right top: Volume, left bottom: demand, right bottom: outflow.

Figure 5-7 shows results of the sensitivity analysis conducted before the calibration of the model. The left graph shows the behaviour of the parameters in the model in which the µ_star shows the influence of the parameters on the model and the σ shows the variability of the influence. The right graph shows the µ_star divided by σ as indication of how good the parameter can be calibrated. It can be concluded that the outflow membership functions are the most sensitive which is in line with logic common sense since the outflow MF's defines the result of the Fuzzy model. it's remarkable that the parameter high outflow is less sensitive than the others but can be declared by the low availability of high outflow data compared to the other categories. The second important parameter are the demand parameters. As raw data analysis gives insight into the behaviour of the outflow which seems to follow some kind of an yearly pattern. The last two important parameters are the high inflow and volume parameters. Those are in line with the assumption of the model being less dependent on the demand when both inflow and volume are high.

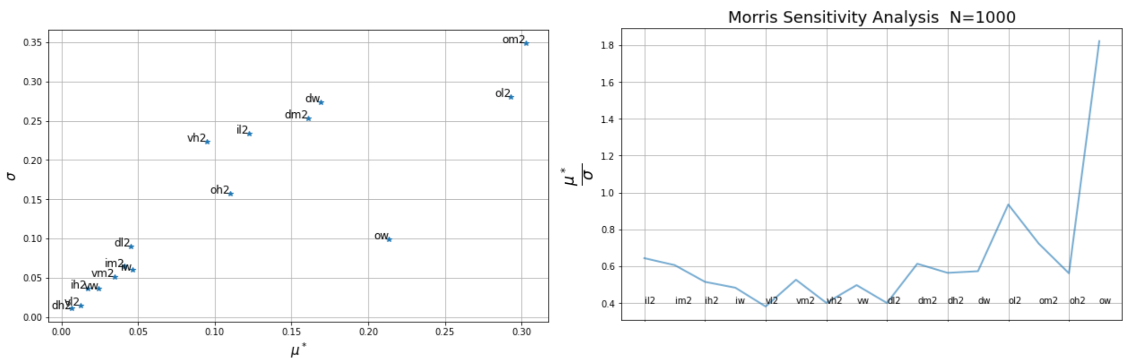


Figure 5-7: The influence and sensitivity of the different parameters. Left shows the influence and variability of the parameter and the right shows the sensitivity of the parameters.

The combination of parameters, giving the highest NSE value as outcome in the sensitivity analysis are used as starting point for the optimization of the model. The parameters with their initial and final value with their relative change are shown in Table 5-6.

Table 5-6: Parameter changes in calibration

PARAMETER	INITIAL VALUE	FINAL VALUE	RELATIVE CHANGE
IL2	165.0	163.4	-1%
VW	818.2	516.4	-37%
DM2	113.1	141.4	25%
DH2	198.4	308.7	51.4%
DW	73.1	59.4	-19%
OL2	8.2	6.1	-25%
OM2	131.6	113.1	-14%
OH2	214.4	183.2	-15%
OW	141.4	124.3	-12%

5.3 Validation

The validation of the applied method has been conducted, revealing that the overall accuracy of the model is lower than anticipated. Nonetheless, the model demonstrates an improvement over the demand and represents progress in the research. Figure 5-8 depicts the moving average of the outflow from the fuzzy model compared to the moving average of the in-situ observations. The key performance indicators (KPIs) provide more statistical insights into the model's performance, as shown in Table 5-6.

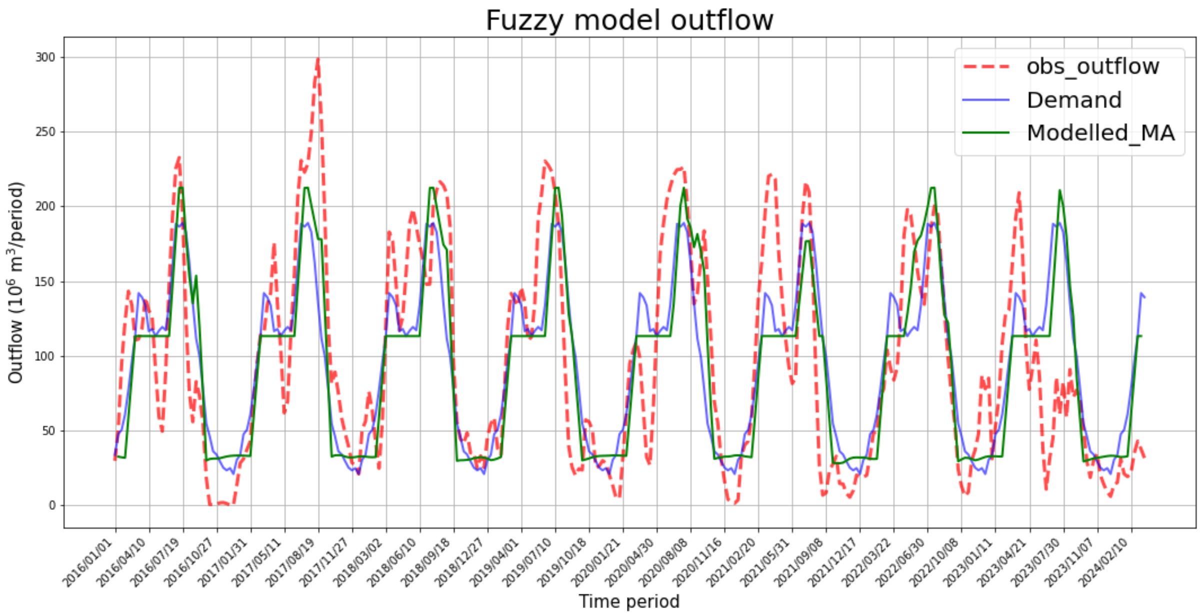


Figure 5-8: Validation of the Fuzzy logic simulated outflow

The moving average of the fuzzy simulation is compared with the moving average of the observations. It can be observed that, on statistical grounds, the model offers a slight improvement over the demand curve. The results, displayed by the KPI's are shown in table 5-6. The first two rows shows the results of the KPI's for box 1 of this method. Giving some insights in the behaviour of the model when no in-situ data is available.

Comparing the accuracy of the simulated volume, on timestep t+1 for 30% of the datapoints, with the overall calculated volume obtained by RS seems to be distributed randomly. Making this method not suitable for being able to say something about the performance of the model. The 2nd method, in which the by fuzzy simulated outflow for the 30% validation dataset is compared with the by water balance obtained outflow for the same 30% of the dataset, gives that the performance is somewhat worse compared to the demand curve compared with the by water balance obtained outflow for the 30% dataset. Although by comparing this with the actual performance of the model, the model is a better representation of the actual outflow than the demand curve. This can be concluded from method 3 in which the 100% of the fuzzy modelled outflow is compared with the observed outflow, and where the demand curve is compared with the observed outflow. This is supported by a higher NS and R2 indicating the pattern is mapped better together with a lower RMSE and MAE which indicates the model has less difference in the variability than the demand curve.

It can be concluded that the remote sensing based model performed better in modelling the outflow of the reservoir compared to the demand curve. Although when no in-situ data is available, the idea can be given that the demand curve performs better.

Table 5-7: Results of the different validation methods demonstrated by the KPI's

	NS	***PBIAS***	***RMSE***	***R2***	***MAE***
(1) Modelled_Volume_Accuracy	0.11	1.77	0.26	0.54	0.18
(2) Model_accuracy_no_in-situ data	0.41	4.08	49.32	0.68	38.04
(2) Demand_accuracy_no_in-situ data	0.41	-0.15	47.72	0.68	35.86
(3) Model_validation	**0.59**	**4.13**	**45.14**	**0.77**	**34.51**
(3) Demand_validation	0.55	0.5	47.29	0.74	35.03

6. Discussion

6.1 Potential

There are several potentials for the application of the methods applied in this research. The potentials are discussed in order of the research questions.

"What is the accuracy of a timeseries containing the volume of the reservoir over the time period from construction till now with a resolution of 10 days average?"
Result section 5.1 shows a high potential in the applied method in obtaining the water volume timeseries of a ungauged dam by showing a NS value of 0.83 and high R^2 value combined with an tendency to overestimate the outflow by 1.93%. The method have a lot of potential in it since only the maximum water level of the reservoir should be know. Making it applicable for locations when there is no in-situ data available by obtaining the maximum water level by optical methods. In addition, combining the two different satellite sources and using both forces is a very valuable method that can be used and further developed in subsequent studies.

"What are the best performing operation rules to simulate the flow in the reservoirs?"
The calibrated fuzzy logic model shows promise for application in ungauged reservoirs where direct measurement data may be scarce or unavailable. In the Fuzzy logic model, lots of parameters can be calibrated. This report has many constrains in the calibration processes of the predefined set of MF's. The sensitivity analysis in section 0 shows that the model is very sensitive to different parameters, especially for the outflow MF. Making it also dependent to what variables are chosen for definition of the MF. The application of the triangular MF's shows reasonable and promising results for further development of the fuzzy logic model in which several other MF's are made.

"What is the accuracy of simulated outflow based on remote sensing data compared with the in-situ data validation?"
The insights gained from assessing the model's accuracy by splitting the remote sensing dataset into calibration and validation sets are not yet reliable for evaluating the model's precision. Section 5.3 demonstrates an accuracy with an NSE of 0.41 and an MAE of 38.04 in the absence of comparison material in the form of in-situ data. However, the actual accuracy of the model is significantly better, with values of 0.59 and 34.51, respectively. This highlights the promise of this method in providing reliable insights when in-situ data is not available. The increase in accuracy can be explained by the application of the moving average on the whole timeseries. However, the validation of this research is not completely independent from the calibration of the model. For both steps, the same datasets are used. The demand curve used as decision variable in the fuzzy logic model is taken as the long term average of the outflow which is the same dataset as is used for validation.

Overall, this research provides a clear understanding of the potential for applying and improving current methodologies to derive operation rules for dams in ungauged areas. The input data for the fuzzy model can be obtained with high accuracy using publicly available remote sensing data. It is demonstrated that fuzzy logic is a promising approach to process this data to derive the operation rules. The operation rules constitute a well-defined set of time-independent rules, allowing for the testing and prediction of specific situations.

6.2 Limitations

Besides the potentials indicated, there are also some limitations to be noted for this research. The first important limitation is the availability of data. Making a reservoir model based on open source data is a challenging task which requires lots of data. Although the amount of data points in this research is limited to 300 values in total. Borregales et al., (2020) emphasized the need for a significant amount of data to calibrate a purely data-driven mode. The method of this research is purely data-driven without a significant amount of data, making the model not suitable for simulating non frequency occurring events, for example the drought in 2023.

Another limitation of this research is the reliance on satellite-based data for calibration and validation phases. Ideally, a model should be calibrated and validated based on reliable, observed data. However, in this study, both the reservoir water volumes and the validation of the fuzzy logic model were based on satellite-based measurements in which error margins are not included. Furthermore, there is no available information regarding the exact measurement methods used. The assumption is that water level and outflow discharge were measured directly, while other values, such as surface area and inflow, were derived from these measurements. The lack of detailed information about these derivations introduces multiple opportunities for errors in the validation system.

Additionally, the location of water level measurements in relation to the outflow can impact the actual measured water levels, potentially leading to discrepancies in the derived water surface area (WSA). As observed in Figure 5-2 and Figure 5-1, the backwater effects from the discharge appear to influence the measured water levels compared to the WSA obtained from satellite data. This discrepancy underscores the potential for measurement errors and highlights the need for more precise and transparent data collection methods to improve the reliability of the model validation process.

Another limitation of this research is about the availability of the WL to WV and WSA curves. For applications inside Vietnam, those curves are available and can be used directly. For further applications in ungauged areas, those curves might be made based on less accurate DEMs or other methods, resulting in more inaccuracies. This research indicates a method to derive the operation rules of a reservoir dam, but uses the known water curves. The influence of the water curves is not tested and so no knowledge of their impact is known.

The last limitation of this research is on the definition and calibration of the membership functions. This research uses a fixed set of MF's for every variable in the Fuzzy model, although the number of MF's can be very sensitive to the outcome of the model. Changing the amount of MF's it is not tested, making the this research limit. The same goes for the calibration of the MF´s. To save computer power and make the calibration easier, constraints are added to the membership functions as well as the range of calibration for them. Consequences of this are less optimal calibrated MF's resulting in a lower accuracy of the fuzzy model it selves.

6.3 Generalisation

Several generalisations on this research van be made. In obtaining the water volumes of the reservoir dam, there are two important steps taken. The first, gathering the WSA based on Sentinel-1 and Sentinel-S2 satellite data and the second, transferring the Sentinel-1 result to model the Sentinel-2 results. The methodology taken can be applied on different reservoirs around the world. Main reason for this is the widely available papers on monthly water surface mapping for nearly all locations in the world. Next to this, there are several ways to estimate the maximum water level for ungauged reservoirs e.g. by analysing DEM maps or google earth. Although it is not analysed in depth but some simple tests showed that the influence of the maximum water level guess is not very sensitive. So it will probably be filtered out by the fusion technics. Making the method applicable for several locations especially in the Red River basin. Application of the obtained operations rules, presented as a set of 27 fuzzy rules can not directly be generalized over other reservoir dams mainly due to the different purposes of the dams. In this research, a dam with main purpose "hydropower" is used and so other decision variables for water subtraction from the reservoir, irrigation demands, flood support factors are not included in the decision making process of the Fuzzy model. However, parts of the methodology can be applied on other dams, such as the water balance to obtain input variables for the fuzzy model, as well as the application of the fuzzy logic with changed input variables. As shown in results section 5.3. Obtaining insight in the accuracy of the model based on the comparison between the observed water volume together with the simulated water volume for t+1 is not feasible. Making it also not generalisable to other locations.

7. Conclusion and recommendations

7.1 Conclusions

This study explored a novel approach to deriving operational rules for reservoir dams in ungauged locations by combining remote sensing data with fuzzy logic modelling. The method proved effective, showing significant promise in developing accurate and reliable operational rules, especially where traditional in-situ data is limited. The calibrated fuzzy logic model successfully enhanced downstream discharge simulations, outperforming traditional methods and highlighting the potential of this integrated approach for improving reservoir management in data-scarce regions.

"What is the accuracy of a timeseries containing the volume of the reservoir over the time period from construction till now with a resolution of 10 days average?"

The results are obtained by the use of the sentinel 1 and sentinel 2 images which achieved high accuracy in estimating the reservoir water volume. The NSE is defined as 0.83 while it has an minor overestimation of volume by 1.93%, indicating the model´s robustness in generating reliable water volume time series in ungauged basins. However there are methods available who achieved a higher accuracy for the sentinel-2 modelling, this research shows the importance and potential of the combination of two satellites in which the strengths of both of them are combined in a final timeseries.

"What is the accuracy of simulated outflow based on remote sensing data compared with the in-situ data validation?"

This study also highlights the possibility of calibration and validation of the model when there is no in-situ data available. The validation of the model by the use of splitting up the remote sensing dataset into a calibration and validation series shows a significant difference with the actual validation with NSE 0.41 for the RS-validation compared with and NSE of 0.59 with the in-situ validation. Although, there is a significant difference. The difference is in a positive direction, the model is performing better than it says its doing.

"What are the best performing operation rules to simulate the flow in the reservoirs?"

The study concludes that the best performing operational rules for simulating reservoir flow are those developed using remote sensing data combined with fuzzy logic modelling. The Ban Chat reservoir's rules, based on inflow, water volume, demand, and outflow, effectively simulated downstream discharge with a Nash-Sutcliffe Efficiency (NSE) of 0.59 and a Mean Absolute Error (MAE) of 34.51—improving accuracy by 7.27% over traditional methods. This research underscores the potential of integrating fuzzy logic with satellite data to create reliable operational rules, particularly in data-scarce regions.

By drawing those conclusions, it should me mentioned that there are a lot of limitations in this research which are bounding the capabilities of the modelling method. By eliminating the limitations as shown in the next – recommendations - section, the performance of the model and so the derivation of the operation rules can improve more.

To finalize, this methodology demonstrated in this research can be applied to derive the operation rules of ungauged reservoirs. The approach's reliance on widely available satellite data and the proven effectiveness of fuzzy logic in modelling reservoir operation rules make it a valuable tool for hydrological studies in data-scarce regions.

7.2 Recommendations

These suggestions are mainly directed towards the VNU-HUS together with other organisations and institutes, working on implementing RS in hydrological modelling of ungauged dams. By adopting these recommendations, the organisations can improve their knowledge about the application of RS with a relatively high temporal scale.

- What can be learned and used from this study?

Combining multiple open source data sources improves the quality as well as quantity of the data.
The use of various publicly available data sources can significantly improve the quality of datasets, especially when in-situ data is unavailable. This study demonstrates that combining the strengths of Sentinel-1 and Sentinel-2 satellites can yield better results. By theoretically evaluating the qualities of each data source—such as the high temporal resolution of Sentinel-1 and the higher spatial resolution of Sentinel-2—a fusion method can be devised that leverages the strengths of both.

Water balance is a sufficient method to obtain the outflow for the fuzzy logic.
Especially in situations where there is no in-situ data available at all. Space observations of the water volume together with by literature proven methods to model the streamflow based on the catchment area such as a VIC-model can be combined by the use of a water balance to obtain a dataset containing the historical downstream discharge of the reservoir dam.

Fuzzy logic is promising but needs further research to improve the quality of the results.
This research shows an easy way to apply fuzzy logic on the water balance to obtain the operation rules of the dam. The model shows some small improvements in comparison to the demand curve.

- What is open for further research?

Model the flow as variation on the demand curve.
Additionally, the setup of the fuzzy logic model warrants further investigation. Future studies should explore whether building a model based on a base flow, such as the demand flow, can yield better results. This involves evaluating different types of membership functions that can be created. To make an informed decision on the number of membership functions, the dataset can be filtered based on a relevant input variable and its associated outflow to assess the feasibility of clustering this outflow. This process might be enhanced through the application of machine learning techniques.

Define a new decision variable to improve the sensitivity of the model.
To enhance the accuracy and reliability of dam operation models, future research should focus on identifying and refining the most influential input parameters, such as including electricity patterns or weather predictions. Conducting on-site interviews with dam operators can provide valuable insights into the decision-making processes and operational behaviours when there is flexibility in dam management. This can help in developing models that better reflect real-world practices by incorporating these behavioural nuances. It is essential to create a comprehensive overview of the decision variables used in the system and understand the basis of operators' decisions, including the interrelationships between different dams and future predictions.

Obtain another source or more insight in the measurements for validation purposes.
Given the potential for significant measurement errors in the observed WSA and WL, it is advisable to explore alternative validation methods. One promising approach is the use of altimetry satellites to directly measure the reservoir's water level. These measurements can then be integrated with the methods proposed in this study to potentially create a more accurate dataset than those obtained from field measurements.

Implementing a hydrological model.

The last recommendation is about the implementation of a hydrological model of the river to predict the inflow of the reservoir dam. This research uses the satellite obtained WSA together with the observed inflow to build the reservoir model. It is expected that the inflow timeseries is a derivation of other measurements without known error margins. By using a hydrological model, for example a VIC-model, which needs to be validated for the specific region before construction of the dam. A better quality of the inflow timeseries can be reached resulting in better, so more clear, paterns in the outflow data. Making better rules which results in a better simulation of the downstream discharge.

References

11761798_08.pdf. (n.d.). Retrieved 26 March 2024, from https://openjicareport.jica.go.jp/pdf/11761798_08.pdf at 2024-05-12

baselinegis. (2021, 9 11). *Satellites Timeline*. Retrieved from BaselineGis: https://baselinegis.com/ retrieved at 2024-04-18

Ali, S. A., & Sridhar, V. (2019). Deriving the Reservoir Conditions for Better Water Resource Management Using Satellite-Based Earth Observations in the Lower Mekong River Basin. *Remote Sensing*, *11*(23), 2872. https://doi.org/10.3390/rs11232872

Associates, M., & Associates (MD&A), M. D. and. (n.d.). *Malcolm Dunstan and Associates—RCC Dam specialists*. RCC Dams - the Home of Malcolm Dunstan and Associates (MD&A). Retrieved 3 July 2024, from https://rccdams.co.uk/dams/ban-chat/

Ban Chat Hydropower Plant. (2021, October 16). https://web.archive.org/web/20211016020042/https://npcetc.vn/en/projects/Hydropower-plants/Ban-Chat-Hydropower-Plant.html

Borregales, M., Møyner, O., Krogstad, S., & Lie, K. (2020). *Data-Driven Models Based on Flow Diagnostics*. *2020*(1), 1–11. https://doi.org/10.3997/2214-4609.202035122

Cardille, J. A., Crowley, M. A., Saah, D., & Clinton, N. E. (Eds.). (2024). *Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications*. Springer International Publishing. https://doi.org/10.1007/978-3-031-26588-4

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., & Li, X. (2016). Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band. *Remote Sensing*, *354*, 1–19. https://doi.org/10.3390/rs8040354

Duc Minh. (*Forthcoming)*, *Working Title of the Report*. Vietnam National University Hanoi University Of Science, *Unpublished Report.*

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, *76*(3), 501–521. https://doi.org/10.1007/BF00939380

Eldardiry, H., & Hossain, F. (2019). Understanding Reservoir Operating Rules in the Transboundary Nile River Basin Using Macroscale Hydrologic Modelling with Satellite Measurements. *Journal of Hydrometeorology*, *20*(11), 2253–2269. https://doi.org/10.1175/JHM-D-19-0058.1

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2). https://doi.org/10.1029/2005RG000183

Fernandes, A., Lima, A., Barbosa, J., & Assis, L. (2019). The influence of spatial discretization on HEC-HMS modelling: A case study. *International Journal of Hydrology*, *3*. https://doi.org/10.15406/ijh.2019.03.00209

Freire, L. (2014). *FUZZY LOGIC APPLIED TO THE MODELING OF WATER DYNAMICS IN AN OXISOL IN NORTHEASTERN BRAZIL (1)*. https://www.semanticscholar.org/paper/FUZZY-LOGIC-APPLIED-TO-THE-MODELING-OF-WATER-IN-AN-Freire/c90cb3628b566bce10c79842f4c7e17d93965ed9

Fundamentals_of_RS_Edited_SC.pdf. (n.d.). Retrieved 27 March 2024, from https://appliedsciences.nasa.gov/sites/default/files/2022-11/Fundamentals_of_RS_Edited_SC.pdf

Gerardo, R., & de Lima, I. P. (2023). Comparing the Capability of Sentinel-2 and Landsat 9 Imagery for Mapping Water and Sandbars in the River Bed of the Lower Tagus River (Portugal). *Remote Sensing*, *15*(7), Article 7. https://doi.org/10.3390/rs15071927

Helmy, B. E.-D. (2023, March 23). *Understanding Otsu's Method for Image Segmentation | Baeldung on Computer Science*. https://www.baeldung.com/cs/otsu-segmentation

Ieva. (2023, May 4). How Many Satellites are in Space? *NanoAvionics*. https://nanoavionics.com/blog/how-many-satellites-are-in-space/

Janssen, J. A. E. B. (2006). *Incorporating qualitative indicators to support river managers; application of fuzzy sets*.

Kebede, M., Wang, L., Yang, K., Chen, D., Li, X., Zeng, T., & Hu, Z. (2020). Discharge Estimates for Ungauged Rivers Flowing over Complex High-Mountainous Regions Based Solely on Remote Sensing-Derived Datasets. *Remote Sensing*, *12*, 1064. https://doi.org/10.3390/rs12071064

Landsat 7 Data Users Handbook | U.S. Geological Survey. (2024). https://www.usgs.gov/landsat-missions/landsat-7-data-users-handbook

Landsat 8 Data Users Handbook | U.S. Geological Survey. (2019). https://www.usgs.gov/landsat-missions/landsat-8-data-users-handbook

Le, D., Le, T. P. Q., Binh, P., Duong, T. T., & Orange, D. (2020). Impact of hydropower dam on total suspended sediment and total organic nitrogen fluxes of the Red River (Vietnam). *Proceedings of the International Association of Hydrological Sciences*, *383*, 367–374. https://doi.org/10.5194/piahs-383-367-2020

Mateo, C. M. R., Hanasaki, N., Komori, D., Yoshimura, K., Kiguchi, M., Champathong, A., Yamazaki, D., Sukhapunnaphan, T., & Oki, T. (n.d.). *A simulation study on modifying reservoir operation rules: Tradeoffs between flood mitigation and water supply*. https://iahs.info/uploads/dms/15613.08-33-40-362-16-H09_2809039_Mateo.pdf

Mateus, P., Catalao, J., & Nico, G. (2017). Sentinel-1 Interferometric SAR Mapping of Precipitable Water Vapor Over a Country-Spanning Area. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(5), 2993–2999. https://doi.org/10.1109/TGRS.2017.2658342

McFEETERS, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

McKone, T. (2005). Can Fuzzy Logic Bring Complex Environmental Problems into Focus? *Environmental Science & Technology*, *39*(2), 42A-47A. https://doi.org/10.1021/es0531632

Mondejar, J. P., & Tongco, A. F. (2019). Near infrared band of Landsat 8 as water index: A case study around Cordova and Lapu-Lapu City, Cebu, Philippines. *Sustainable Environment Research*, *29*(1), 16. https://doi.org/10.1186/s42834-019-0016-5

Mujumdar, P. P., & Ghosh, S. (2008). Fuzzy Logic-Based Approaches in Water Resource System Modelling. In R. J. Abrahart, L. M. See, & D. P. Solomatine (Eds.), *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications* (pp. 165–176). Springer. https://doi.org/10.1007/978-3-540-79881-1_12

Nikoo, M. R., Karimi, A., Kerachian, R., Poorsepahy-Samian, H., & Daneshmand, F. (2013). Rules for Optimal Operation of Reservoir-River-Groundwater Systems Considering Water Quality Targets: Application of M5P Model. *Water Resources Management*, *27*(8), 2771–2784. https://doi.org/10.1007/s11269-013-0314-3

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. IEEE Transactions on Systems, Man, and Cybernetics. https://doi.org/10.1109/TSMC.1979.4310076

Peña-Luque, S., Ferrant, S., Cordeiro, M. C. R., Ledauphin, T., Maxant, J., & Martinez, J.-M. (2021). Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level. *Remote Sensing*, *13*(16), 3279. https://doi.org/10.3390/rs13163279

Pham, L. (2021, June 2). *How China's Upstream Dams Affect Communities Along the Red River*. https://saigoneer.com/saigon-environment/25217-how-china-s-upstream-dams-affect-communities-along-the-red-river

Pham, Q.-N., Nguyen, N.-H., Ta, T.-T., & Tran, T.-L. (2023). Vietnam's Water Resources: Current Status, Challenges, and Security Perspective. *Sustainability*, *15*(8), 6441. https://doi.org/10.3390/su15086441

Prime minister of Vietnam. (2019, June 17). *Decission Regarding the promulgation of inter-reservoir operating procedures in the Red River basin*. (In Vietnamees)

Qiu, F., Berglund, J., Jensen, J., Thakkar, P., & Ren, D. (2004). Speckle Noise Reduction in SAR Imagery Using a Local Adaptive Median Filter. *Giscience & Remote Sensing - GISCI REMOTE SENS*, *41*, 244–266. https://doi.org/10.2747/1548-1603.41.3.244

Sathianarayanan, M., Saraswat, A., Athick, A. S. M. A., & Lin, H.-M. (2023). Intercomparison between sentinel-1, sentinel-2, and landsat-8 on reservoir water level estimation. *Sustainable Water Resources Management*, *9*(6), 185. https://doi.org/10.1007/s40899-023-00974-4

Sentinel-2 Products Specification Document. (2012). https://sentinel.esa.int/documents/247904/685211/S2-PDGS-TAS-DI-PSD-V14.9.pdf

Shen, G., & Fu, W. (2020). Water Body Extraction using GF-3 Polsar Data—A Case Study in Poyang Lake. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 4762–4765. https://doi.org/10.1109/IGARSS39084.2020.9323878

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, *523*, 739–757. https://doi.org/10.1016/j.jhydrol.2015.02.013

Souza, W. D. O., Reis, L. G. D. M., Ruiz-Armenteros, A. M., Veleda, D., Ribeiro Neto, A., Fragoso Jr., C. R., Cabral, J. J. D. S. P., & Montenegro, S. M. G. L. (2022). Analysis of Environmental and Atmospheric Influences in the Use of SAR and Optical Imagery from Sentinel-1, Landsat-8, and Sentinel-2 in the Operational Monitoring of Reservoir Water Level. *Remote Sensing*, *14*(9), 2218. https://doi.org/10.3390/rs14092218

THE RED RIVER BASIN ORGANIZATION IN INTEGRATED WATER RESOURCES MANAGEMENT OF VIET NAM. (n.d.). Retrieved 26 March 2024, from https://www.narbo.jp/data/01_events/materials/sawf_01_2-3.pdf

Thương B. C. (2023, December 31). *Ngày này năm xưa 31/12: Khánh thành Nhà máy Thủy điện Bản Chát | Báo Công Thương*. Báo Công Thương điện tử, kinh tế, chính trị, xã hội.

https://congthuong.vn/ngay-nay-nam-xua-3112-khanh-thanh-nha-may-thuy-dien-ban-chat-295084.html

Tone, S. (2023, October 12). *China, Mekong Countries One Step Closer to Real-Time Water Data Sharing*. #SixthTone. https://www.sixthtone.com/news/1013871

Wang, Y., Chang, J., & Huang, Q. (2010). Simulation with RBF Neural Network Model for Reservoir Operation Rules. *Water Resources Management*, *24*(11), 2597–2610. https://doi.org/10.1007/s11269-009-9569-0

Water-vital-vietnam-future.pdf. (n.d.). Retrieved 25 April 2024, from https://www.adb.org/sites/default/files/publication/29806/water-vital-vietnam-future.pdf

Winkel, W. (2022). *An impact assessment of human interventions on the hydrological regime of the black river basin in Vietnam.* (Bachelor Thesis 1; p. 51). https://essay.utwente.nl/93485/1/Winkel-Walter.pdf

Wu, J., & Zhu, X. (2006). Using the Shuffled Complex Evolution Global Optimization Method to Solve Groundwater Management Models. In X. Zhou, J. Li, H. T. Shen, M. Kitsuregawa, & Y. Zhang (Eds.), *Frontiers of WWW Research and Development—APWeb 2006* (pp. 986–995). Springer. https://doi.org/10.1007/11610113_105

Appendices

Appendix A

As explained in section 4.3.3,in the sensitivity analysis is conducted in a certain range for the parameters. The ranges of the parameters are defined per parameter and are shown in table A-1. These boundaries are set, based on the observed datasets for the reservoir dam. So minimum observed inflow means the lowest inflow which can be found in the dataset. And the range is the distance between the lowest and the highest value. See figure A-1 for clarification in an image.

Tabel A-0-1: Boundary conditions Morris sensitivity analysis

il2	[Minimum observed inflow, Minimum observed inflow + 1/3 inflow_range]
im2	[Minimum observed inflow + 1/3 inflow_range, Minimum observed inflow + 2/3 inflow_range]
ih2	[Minimum observed inflow + 2/3 inflow_range, Maximum observed inflow]
iw	[inflow_range/2, inflow_range]
vl2	[Minimum observed volume, Minimum observed volume + 1/3 volume_range]
vm2	[Minimum observed volume + 1/3 volume_range, Minimum observed volume + 2/3 volume_range]
vh2	[Minimum observed volume + 2/3 volume_range, Maximum observed volume]
vw	[volume_range/2, volume_range]
dl2	[Minimum observed demand, Minimum observed demand + 1/3 demand_range]
dm2	[Minimum observed demand + 1/3 demand_range, Minimum observed demand + 2/3 demand_range]
dh2	[Minimum observed demand + 2/3 demand_range, Maximum observed demand]
dw	[demand_range/2, demand_range]
ol2	[Minimum observed outflow, Minimum observed outflow + 1/3 outflow_range]
om2	[Minimum observed outflow + 1/3 outflow_range, Minimum observed outflow + 2/3 outflow_range]
oh2	[Minimum observed outflow + 2/3 outflow_range, Maximum observed outflow]
ow	[outflow_range/2, outflow_range]

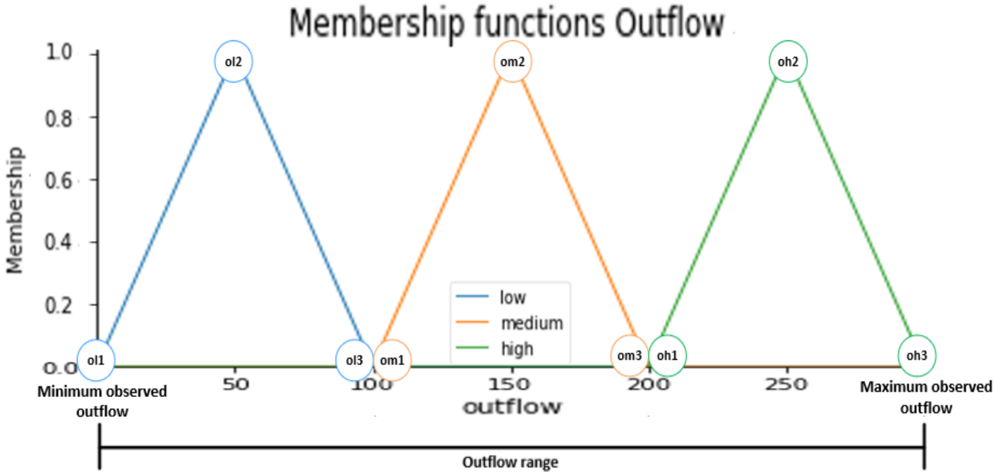


Figure A-0-1: clarification for the lowest, highest and range for the sensitivity analysis

Appendix B

In Table B-1, an overview is provided of how the coordinates of the membership functions are determined in the Python code. For example the il2 is a set value after the optimisation and lower and upper boundaries are made out of them by the use of the width of the MF and some constrains to make the code work.

Tabel B-1: Coordinates for the final MF's

il1	0
il2	il2
il3	il2 + 1/2 * iw
im1	min(il3, max(0,im2 - 1/2 * iw))
im2	im2
im3	im2 + 1/2 * iw
ih1	min(im3,ih2 - 1/2 * iw)
ih2	ih2
ih3	max(ih2 + 1/2 * iw, statistics_matrix.loc["Maximum","inflow"])
vl1	0 #max(0, vl2 - 1/2 * vw)
vl2	vl2
vl3	vl2 + 1/2 * vw
vm1	min(vl3, max(0,vm2 - 1/2 * vw))
vm2	vm2
vm3	vm2 + 1/2 * vw
vh1	min(vm3,vh2 - 1/2 * vw)
vh2	vh2
vh3	max(vh2 + 1/2 * vw, statistics_matrix.loc["Maximum","volume"])
dl1	0 #max(0, dl2 - 1/2 * dw)
dl2	dl2
dl3	dl2 + 1/2 * dw
dm1	min(dl3, max(0,dm2 - 1/2 * dw))
dm2	dm2
dm3	dm2 + 1/2 * dw
dh1	min(dm3,dh2 - 1/2 * dw)
dh2	dh2
dh3	max(dh2 + 1/2 * dw, statistics_matrix.loc["Maximum","demand"])
ol1	0 #max(0, ol2 - 1/2 * ow)
ol2	ol2
ol3	ol2 + 1/2 * ow
om1	min(ol3, max(0,om2 - 1/2 * ow))
om2	om2
om3	om2 + 1/2 * ow
oh1	min(om3,oh2 - 1/2 * ow)
oh2	oh2
oh3	max(oh2 + 1/2 * ow, statistics_matrix.loc["Maximum","MA_sim_outflow"])
#in	w membership functions
il1	0
il2	il2
il3	il2 + 1/2 * iw
im1	min(il3, max(0,im2 - 1/2 * iw))
im2	im2
im3	im2 + 1/2 * iw
ih1	min(im3,ih2 - 1/2 * iw)
ih2	ih2